Effects of a Strategy-Focused Instructional Program on the Writing Quality of Upper Elementary Students in the Netherlands

Renske Bouwer, Monica Koster, and Huub van den Bergh
Utrecht University

In this study, the authors tested the effects of Tekster [Texter], a comprehensive strategy-focused writing instruction program, using a switching replication design with three measurement occassions. The program was implemented by fourth, fifth, and sixth grade teachers ($N = 76$) in 60 general education classrooms in the Netherlands. Students ($n = 688$) and teachers ($n = 31$) in Group 1 worked with Tekster during the first 8-week period, between the first and second measurement occasion. Students ($n = 732$) and teachers ($n = 45$) in Group 2 implemented Tekster during the second 8-week period, between the second and third measurement occasion. The intervention led to statistically significant improvements in the quality of students' writing. The effect size for the full sample was 0.32 and 0.40 for students who students who completed all 16 Tekster lessons. Gains shown by students in Group 1 were maintained after 8 weeks. Because writing quality was assessed in 3 genres, the findings are generalizable across students, classes, and writing tasks. Taken together, the results of this study demonstrate that a strategy-focused writing instruction program, such as Tekster, can be an effective way to improve upper-elementary students' written language skills.

Educational Impact and Implications Statement
This study shows how Tekster [Texter], a strategy-focused writing instruction program, improves the writing performance of students in Grade 4 to 6. This positive effect was still visible 2 months after the intervention. As the intervention was successfully implemented by teachers in a large number of classrooms, this study suggests that Tekster is a promising approach for improving students' writing in general education.

Keywords: writing, writing instruction, observational learning, strategy instruction, elementary grades

Supplemental materials: http://dx.doi.org/10.1037/edu0000206.supp

Despite the fact that writing plays an important role in academic and career success, research shows large numbers of students from many different countries fail to develop essential writing skills (e.g., Department for Education, 2012; Salahu-Din, Persky, & Miller, 2008). For example, a recent national assessment in the Netherlands revealed most elementary-aged students were unable to write texts that convey a single, simple message to the reader and students' writing skills improved negligibly from fourth to sixth grade (Kuhlemeier, Til, Hemker, de Klijn, & Feenstra, 2013). Furthermore, the Dutch Inspectorate of Education determined the quality of writing instruction to be sufficient in only one third of the nation's schools (Henkens, 2010). Thus, an improvement in elementary-level writing instruction in the Netherlands is required. For this purpose, we developed the writing program Tekster. Tekster incorporates several research-supported instructional practices and addresses both the focus and mode of instruction (what we teach and how we teach it). The effectiveness of Tekster for fourth-, fifth-, and sixth-grade Dutch students was tested in this study.

Focus of Instruction

The major problem developing writers face is cognitive overload. Writers have to perform several resource-demanding cognitive activities simultaneously, such as activating prior knowledge, generating content, planning, formulating, and revising—all while taking into account the communicative goal of the text and the intended audience (Fayol, 1999). The amount of attention required for foundational skills (e.g., handwriting, spelling, sentence, and paragraph construction) also needs to be considered with developing writers because they often lack automaticity in these areas (McCutchen, 2011). Developing writers predominantly use a "knowledge-telling" approach to overcome cognitive overload. That is, they write whatever happens to come to mind and typically

focus only on the content of their texts (Bereiter, Burtis, & Scardamalia, 1988). With this approach text production is restrained by the storage and retrieval capacity of short-term memory (STM; Miller, 1956) and this often results in texts that are not sufficiently adapted to the communicative goal and intended audience (Berninger et al., 1992; McCutchen, 1996). To improve students' writing performance, instruction should be aimed at helping them develop the knowledge and skills required to manage the cognitive overload that often occurs when composing.

Strategy Instruction

An effective way to help developing writers manage cognitive overload is to teach them to use strategies that reduce the number of cognitive processes that are active at the same time (Kellogg, 1988, 2008). For instance, when students are taught to plan during the prewriting phase, they can focus on other processes while drafting. A substantial body of research has examined the impact of explicitly teaching students to use writing strategies. Some studies investigated strategies designed to guide general writing processes, such as brainstorming (Troia & Graham, 2002) or revising (Fitzgerald & Markham, 1987), whereas others featured genre or task specific strategies, such as writing a narrative text (Brunstein & Glaser, 2011) or a persuasive essay (Wong, Hoskyn, Jai, Ellis, & Watson, 2008). Despite the diversity of research examining explicit strategy instruction, results are remarkably consistent and positive. For example, several recent meta-analyses reported large average weighted effect sizes (ESs), ranging from 0.82 to 1.15, for explicit strategy instruction (Graham, 2006; Graham, McKeown, Kiuhara, & Harris, 2012; Graham & Perin, 2007; Hillocks, 1984; Koster, Tribushinina, De Jong, & Van den Bergh, 2015).

Self-Regulation

When explicit strategy instruction is combined with teaching self-regulatory skills, the impact on students' writing is even greater (Graham et al., 2012). *Self-regulation* is "the process whereby individuals activate and sustain behaviors, cognitions, and affect, which are systematically oriented toward the attainment of goals" (Schunk, 2012, p. 123). Essential self-regulatory skills in writing include setting communicative, process, and progress goals, and subsequently monitoring progress toward those goals (Flower & Hayes, 1981). The most prominent and well-researched model for explicitly teaching writing strategies and self-regulation is self-regulated strategy development (SRSD; Harris, Graham, Mason, & Saddler, 2002). SRSD has been validated through research spanning over three decades and involving a wide range of students in many different instructional environments. Results from SRSD studies consistently show the approach is highly effective for improving students' writing performance (ES = 1.17, Graham et al., 2012).

Students' self-regulation is positively affected by the attainment of specific goals which, in turn, enhances self-efficacy for writing (Latham & Locke, 1991; Schunk, 1990). Students benefit the most from challenging, but attainable, goals that specify what needs to be accomplished through the writing task (Schunk, 1990). For example, previous research shows assigning students specific goals for improving the content of their texts and making them aware of the intended audience leads to improvements in planning, drafting, and revising (Ferretti, Lewis, & Andrews-Weckerly, 2009; Ferretti, MacArthur, & Dowdy, 2000; Graham, MacArthur, & Schwartz, 1995; Midgette, Haria, & MacArthur, 2008). Research also indicates short-term writing goals are more beneficial than goals spanning longer periods of time (Latham & Locke, 1991).

Text Structure Instruction

To be proficient writers, students need to be able to establish their own composing goals for different writing tasks. They also need to know how to create texts that meet the goals they set (Schoonen & De Glopper, 1996). Explicit text structure instruction, whereby the elements and organization of different text types are specifically taught, has been shown to help students acquire the knowledge needed to set and achieve writing goals. Research examining the impact of explicit text structure instruction for elementary-aged students spans three major genres: narrative (Fitzgerald & Teasley, 1986; Gordon & Braun, 1986), persuasive (Crowhurst, 1990, 1991; Scardamalia & Paris, 1985), and informative (Bean & Steenwyk, 1984; Raphael & Kirschner, 1985). The findings from two recent meta-analyses provide further support for the positive effect of text structure instruction. Graham et al. (2012) and Koster et al. (2015) reported an average weighted ES for explicit text structure instruction of 0.59 and 0.76, respectively.

Mode of Instruction

For developing writers, learning to write and task execution are often inextricably linked. Simultaneously, students have to learn how to write and produce texts (Rijlaarsdam & Couzijn, 2000). However, because text production is so cognitively demanding for developing writers, this instructional approach often results in students having minimal attentional capacity left to learn from their writing experiences (Rijlaarsdam & Couzijn, 2000). Thus, to optimize the way writing is taught, it is important to carefully consider the format and sequence of instruction.

Observational Learning

One way to separate task performance from learning is to provide opportunities for observation (Zimmerman & Risemberg, 1997). Observing someone complete an unfamiliar task is less demanding on working memory than having to actually perform the task yourself. This is particularly true when the skill being learned is cognitively complex—such as writing (Rijlaarsdam, 2005). Observational learning was first described and studied by Bandura (1986) as part of social–cognitive learning theory. Within this framework, observation allows individuals to gain insight into the usefulness and consequences of the behavior being modeled. Behavior that is evaluated positively and considered useful will be retained (Schunk, 2012). Observational learning can be applied to teaching writing in two ways: through different types of modeling (before and during writing) and through reader feedback (during and after writing).

Teacher modeling. In writing instruction, observational learning is frequently implemented by means of teacher modeling.

Modeling involves explaining, demonstrating, and verbalizing one's thoughts and actions, with the aim of eliciting behavioral change in an observer (Schunk, 2012). This kind of modeling prepares students for the forthcoming composing task in the initial phase of the writing process. Several studies have demonstrated the effectiveness of teacher modeling as an instructional practice for teaching writing strategies (e.g., Fidalgo, Torrance, Rijlaarsdam, Van den Bergh, & Lourdes Álvarez, 2015; Graham, Harris, & Mason, 2005).

Mastery versus coping models.   Models can show either mastery or coping behavior. Mastery models demonstrate a flawless performance, whereas coping models display common challenges, as well as ways to overcome those difficulties and gradually improve performance (Schunk, 1987; Zimmerman & Kitsantas, 2002). In a study on revision skills, Zimmerman and Kitsantas (2002) found observing a coping model raised students' self-efficacy and enhanced their performance more effectively than a mastery model. Research suggests observing coping models is particularly beneficial for weaker students. This may be the result of explicitly seeing how to overcome difficulties and/or watching someone who is perceived as similar improve performance over time (Schunk, 1987).

Peer modeling.   When peers—rather than teachers—act as models, perceived model-observer similarity is even higher because of the developmental resemblance (Schunk, 1987). Peer modeling has been investigated in several studies. Raedts, Rijlaarsdam, Van Waes, and Daems (2007) found observing video-based peer models improved text organization and self-perception of writing performance. Couzijn (1999) demonstrated observing peer models can have large effects on argumentative text-writing. Van Steendam, Rijlaarsdam, Van den Bergh, and Sercu (2014) found both more and less proficient writers benefited from peer modeling during a collaborative revising task. Braaksma (2002) and Braaksma, Rijlaarsdam, Van den Bergh, and Van Hout-Wolters (2004) found observing peer models positively impacted students' writing performance and writing processes. Braaksma's (2002) findings also provided support for the model-observer similarity hypothesis. Weaker students performed better after focusing on a weaker peer model, whereas stronger students showed greater improvement after focusing on a stronger peer model. Observing mastery peer models may be especially beneficial for stronger students because they set positive standards for performance (Zimmerman & Kitsantas, 2002). In contrast, observing coping peer models may be especially effective for weaker students, as they enhance self-efficacy and motivation (Schunk, 1987). It should be noted, however, all the aforementioned peer modeling studies were conducted with (post)secondary students, rather than students in the elementary grades.

Reader reaction.   Whereas teacher and peer modeling primarily focus on teaching students aspects of the writing process, a different form of observational learning can be used to provide students with feedback on the communicative effectiveness of their compositions. In contrast with oral communication, separation in time and space results in writers rarely receiving any direct cues or feedback from those who read their text (Rijlaarsdam et al., 2008). This can be particularly disadvantageous for developing writers who are not yet proficient in self-evaluation. Observational learning can help bridge this gap and develop students' understanding of how readers experience and perceive their texts (Couzijn & Rijlaarsdam, 2004; Schriver, 1992). Several researchers (Couzijn, 1995; Couzijn & Rijlaarsdam, 2004; Holliway & McCutchen, 2004; Rijlaarsdam, Couzijn, Janssen, Braaksma, & Kieft, 2006) have shown students' writing can improve after observing the effect their text has on readers. Meta-analytic results indicate both feedback and peer interaction can enhance writing quality. The average weighted ESs reported by Graham et al. (2012) and Koster et al. (2015) for feedback were 0.80 and 0.88, and for peer interaction were 0.89 and 0.59, respectively.

Gradual Release of Responsibility

Improving students' writing performance cannot be accomplished solely through observational learning; there comes a time when students need to transition from observing writing models to actually composing themselves. Moreover, to successfully complete a writing task, students must eventually progress through all the stages of the writing process. One way to ease the transition between observation and task execution is through the gradual release of responsibility (Pearson & Gallagher, 1983). With this approach, cognitive load is gradually shifted from observing models, to guided practice, and finally to independent performance. The gradual release of responsibility model builds on Vygotsky's (1980) sociocultural theory and concept of the zone of proximal development. Vygotsky defined the zone of proximal development as the area between a student's level of independent performance and potential development, as determined by assisted performance. Teachers can facilitate progression from assisted to independent performance through scaffolding. That is, they control elements of a task initially beyond a student's capacity to enable the development of skills within the range of competence (Wood, Bruner, & Ross, 1976). As a student progresses, teacher assistance is gradually reduced. For scaffolding to be successful, teachers need to help students develop strategies that are transferrable to new tasks and situations (Bodrova & Leong, 1998).

Writing instruction programs that use gradual release of responsibility and scaffolding techniques have been shown to improve students' written language skills (Graham et al., 2005; Graham et al., 1995). Many of these programs also use explicit instruction to activate students' background knowledge and help them understand the purpose and benefits of the strategy being taught. For upper-elementary aged students, generalization of strategy use to other tasks and domains is promoted through comprehensive and explicit instruction regarding how and when a strategy can best be applied (O'Sullivan & Pressley, 1984).

Aim of the Study

The main purpose of this study was to test the effectiveness of Tekster (Koster, Bouwer, & Van den Bergh, 2014a, 2014b, 2014c), a comprehensive writing instruction program we developed to be implemented by fourth, fifth, and sixth grade general education teachers in the Netherlands. The main focus of Tekster is teaching students a general writing strategy, as well as the self-regulation skills needed to use the strategy successfully. Genre-specific features are addressed through explicit instruction in text structure. The predominant mode of instruction is observational learning, complemented by explicit instruction and guided practice that includes extensive scaffolding and the gradual release

of responsibility (Wood et al., 1976). In this regard, Tekster bears close resemblance to SRSD (Harris et al., 2002) and cognitive self-regulation instruction (Fidalgo et al., 2015).

In the present study, we investigated whether Tekster improved the quality of writing produced by fourth, fifth, and sixth grade Dutch students and whether the effect of the intervention was maintained over time. In addition, we examined whether the effect differed based on students' grade level, gender, or level of writing proficiency.

Method

Sample

Seventy-six upper-elementary teachers, representing 60 classrooms, volunteered to participate in the study. The majority of teachers were female (82%) and all participants held the required professional certification. The study took place in 27 schools, located throughout the Netherlands. Eleven schools were in the northern part of the country, nine were centrally located, and seven were in the southern region. Sixty percent of the schools were religiously affiliated (11 Catholic, two Protestant, two Reformed, one Islamic) and 40% were public. Ten schools had one participating classroom, whereas two to five classrooms participated in the other 17 schools. With regard to grade level, there were 20 fourth-grade classes, 13 fifth-grade classes, 16 sixth-grade classes, and 11 multigrade classes (i.e., a combination of two or three grade levels). The average number of students per class was 23.6 (SD = 5.6), half of whom were female. The schools, teachers, and students in our sample did not differ significantly from the Dutch population in terms of denomination (Ministry of Education, Culture, & Science, 2015), gender (Central Office for Statistics, 2015; Inspectorate of Education, 2012), or classroom size (Central Office for Statistics, 2015).

In total, 1,420 students participated in the study: 477 fourth graders (M age = 9.40, SD = 0.62), 454 fifth graders (M age = 10.40, SD = 0.61), and 489 sixth graders (M age = 11.50, SD = 0.64).[1] A small number of individual students dropped out because they changed schools during the study. Specifically, 17 students (1.2%) completed only the pretest measures and 37 students (2.6%) completed only one of the two posttest measures.

Design of the Study

To analyze whether Tekster improved students' writing quality, we used a switching replication design (Shadish, Cook, & Campbell, 2002) with two groups and three measurement occasions (M1, M2, M3; see Table 1). In the first phase of the study, from M1 to M2, teachers and students in Group 1 worked with Tekster—instead of their regular writing instruction program—for 8 weeks, completing two lessons per week. Group 2 served as a control group during this period; teachers and students continued with their existing writing activities and routines. During the second 8-week phase, between M2 and M3, the intervention switched between groups. Group 2 implemented Tekster and Group 1 returned to their original writing program. M3 served as a posttest for students in Group 2, as well as a delayed posttest for students in Group 1, which enabled us to measure their level of retention.

A switching replication design is superior to a regular pretest posttest (quasi-) experimental design because the intervention is implemented in both groups, but during different time intervals (Shadish et al., 2002). It is not only a more ethical design, as all students eventually benefit from the intervention, but it also allows for a test of internal validity. If the intervention is equally effective in both groups, the effect does not likely depend on characteristics of a particular group. If the effect of the intervention is not equally effective in both groups, internal validity might be threatened. Moreover, because the intervention is replicated in two groups, important information about the reproducibility and generalizability of the results is generated (Open Science Collaboration, 2015). The design also provides information about maintenance effects because it includes a delayed posttest (M3) for students in Group 1.

Assignment of schools to groups. The school holiday calendar determined which schools were assigned to Group 1 and Group 2. Specifically, schools located in the northern region were assigned to Group 1 and those in the south were assigned to Group 2. Schools from the middle region were randomly assigned to Group 1 or 2. Group 1 included 14 schools, 31 teachers (84% female), and 29 classes. Group 2 included 13 schools, 45 teachers (80% female), and 31 classes. Table 2 contains a summary of student information for each group. The number of students per grade was similar for both groups, $\chi^2(2) = 2.67$, $p = .26$, and there were no statistically significant differences in gender distribution, $\chi^2(1) = 2.21$, $p = .14$, or age, $t(1414) = -1.31$, $p = .19$, between groups.

Writing Instruction

Existing instruction. The intervention program was compared to the existing writing instruction practices used in each participating classroom. In the Netherlands, writing is traditionally taught as part of the Dutch language curriculum. According to a report published by the Dutch Inspectorate of Education (Henkens, 2010), of the 8 hours per week reserved for language teaching, an average of only 45 min is devoted to writing. Writing lessons are primarily product-focused: Students receive minimal support during the writing process and are not taught how to approach writing tasks. In addition, in the majority of schools, students' writing performance is not monitored and they are rarely given feedback on their compositions. Many of the Inspectorate's findings were recently corroborated by a study exploring how writing is taught by 51 Dutch elementary teachers (Rietdijk, Van Weijen, Janssen, Van den Bergh, & Rijlaarsdam, 2015). For example, 94% of the teachers said they spend less than one hour per week teaching writing. They also described typical writing lessons as dominated by independent student work, with only one third of the time being used for plenary instruction. Modeling, individualized support, and providing students with feedback were all reported to be uncommon practices. In contrast with what was reported by the Inspectorate, however, teachers who participated in Rietdijk et al.'s study

---

[1] Specific information on students' special educational needs was not available. Typically, in an average Dutch general education classroom, 20% to 25% of the students will have learning and/or behavioral difficulties that require additional, individualized attention (Koopman, Ledoux, Karssen, Van der Meijden, & Petit, 2015).

Table 1
Design of the Study

Group	Pretest (M1)	Phase 1 (8 weeks)	Posttest (M2)	Phase 2 (8 weeks)	Delayed posttest (M3)
1	Tasks a, b, c	Tekster intervention	Tasks d, e, f	Existing writing instruction	Tasks g, h, i
2		Existing writing instruction		Tekster intervention	

said they do attend to the different stages of the writing process. For example, a majority of respondents reported using prewriting activities and half said they ask students to revise their texts.

Tekster. The intervention program, Tekster, included a series of 16 grade-level specific lessons, compiled in a student workbook and accompanied by a teacher's manual (Koster et al., 2014a, 2014b, 2014c). Tekster incorporates several different research-based practices to address both the focus and the mode of instruction. Table 3 gives an overview of how the program's three design principles—writing strategies, text structure, and self-regulation skills (see Rijlaarsdam, Janssen, Rietdijk, & Van Weijen, in press)—were operationalized into specific teaching and learning activities.

Lesson format and writing strategies. Tekster lessons followed a generally consistent format, with each lesson typically lasting between 45 and 60 min (see Table 4). The focal point of instruction was the writing strategy students learned to help guide them through the steps of the writing process. A mnemonic device was used to help students remember and apply the writing strategy: The first letter of each strategy step formed an acronym that spelled the name of an animal.

- Grade 4 students learned VOS *(which means fox)*: Verzinnen *(generate content)*, Ordenen *(organize)*, Schrijven *(write)*.
- Grade 5 students learned DODO (which means *dodo*): Denken *(think)*, Ordenen *(organize)*, Doen *(do)*, Overlezen *(read)*.
- Grade 6 students learned EKSTER (which means *magpie*): Eerst nadenken *(think first)*, Kiezen & ordenen *(choose & organize)*, Schrijven *(write)*, Teruglezen *(reread)*, Evalueren *(evaluate)*, Reviseren *(revise)*.

The three animals were used as a common theme for all the lessons in the corresponding grade level and small images representing the animals provided additional visual support. A sample Tekster lesson is included as Supplemental Appendix A in the online supplemental materials.

Lesson content and sequence. During the first Tekster lesson, students were introduced to the acronym animal corresponding with the writing strategy they would learn, through a story. They also practiced the steps of the strategy for the first time. In subsequent lessons, students learned to apply the writing strategy to different types of texts. All the practice writing tasks were authentic and represented a variety of communicative goals and audiences. For instance, students in each grade wrote texts that were descriptive (e.g., personal advertisement, self-portrait), narrative (e.g., story for kindergartener, newspaper article), persuasive (e.g., email nominating for a TV program, flyer recruiting new members for a club), instructive (e.g., recipe, rules for a game) and personal communications (e.g., holiday postcard, party invitation). The writing tasks for each grade level were in line with the Dutch Ministry of Education's goal for students at the end of elementary school "to write coherent texts, with a simple linear structure on various familiar topics; the text includes an introduction, body, and ending" (Expert Group Learning Trajectories, 2009, p.15).

The level of difficulty for the writing tasks ascended through the grades as follows: In Grade 4, tasks featured an intended audience in close proximity to the student, such as classmates, friends, and (grand)parents. In Grade 5, the target audience expanded to include people with whom students had a more distal relationship but yet were still familiar, such as teachers, relatives, and neighbors. In Grade 6, students also wrote texts intended for unfamiliar people, such as a newspaper editor and owner of a company.

Lesson development. Tekster lessons were developed in close collaboration with 16 elementary school teachers. These teachers were divided into three design teams that met once a month over a period of six months. After receiving an introduction to the program's guiding principles, two design teams worked on developing the practice writing tasks that would eventually be integrated into Tekster lessons. Each writing task needed to focus on a topic of interest to upper-elementary students and have a clearly specified communicative goal and target audience. Teachers piloted the writing tasks with their own students and received feedback from their team members and the authors during the monthly meetings. The third design team made peer modeling video clips that were used as part of Tekster instruction. After the writing tasks and video clips were created, the authors wrote the detailed lesson plans for each grade level and subsequently piloted the program (see Koster, Bouwer, & Van den Bergh, 2016).

Teacher training. The teachers who participated in this study learned about Tekster during a 4-hr session training session led by the authors and held in small groups consisting of no more than 12 people. At the beginning of the training session, each teacher received a Tekster teacher's manual that was divided into two sections. The first section included an introduction to the program (e.g., goals, guiding principles) and descriptions of the essential components (e.g., instructional model, general lesson format and sequence, specific research-based practices). An overview of the

Table 2
Student Characteristics

	Group 1			Group 2		
Grade	n	% female	Mean age (SD)	n	% female	Mean age (SD)
4	245	47%	9.41 (.58)	232	54%	9.39 (.65)
5	217	51%	10.39 (.63)	237	54%	10.42 (.59)
6	226	46%	11.50 (.67)	263	48%	11.50 (.62)
Total	688	48%	10.41 (1.07)	732	52%	10.48 (1.07)

Table 3

Overview of Design Principles, Learning and Teaching Activities of Tekster

Design principles		Tekster intervention	
Focus of instruction	Mode of instruction	Learning activities	Teaching activities
1. Writing strategies	a. Observational learning	Observe/discuss/compare model(s) (teacher or peer) and apply the writing strategy in different stages of the writing process	Model strategy use by thinking aloud while performing (part of) the writing task
	b. Explicit instruction	Listen actively, retrieve relevant background knowledge from memory, take notes	Explain the components of the strategy, make students aware of the purpose and benefits of using a writing strategy, activate students' background knowledge
	c. (Guided) practice	Apply the steps of the strategy to authentic writing tasks in various genres with clear communicative goals and intended audience	Provide help when needed through scaffolding and process feedback
2. Text structures	a. Observational learning	Before writing: Observe/discuss/ compare model(s), (teacher or peer) talking about criteria for various text types, compare and discuss model texts of the same text type to derive criteria and conventions for a good text	Before writing: Model the relevant aspects of the text type, provide model texts or show video clips of peer modeling
		After writing: Evaluate peer/own text on the basis of the previously discussed criteria and give feedback (reader reaction), observe reader reaction, observe model revising on the basis of feedback	After writing: Evaluate students' texts on the basis of previously discussed criteria, give feedback (reader reaction), model how to revise the text
	b. Explicit instruction	Listen actively, take notes	Explain why and how the criteria and conventions should be used, discuss important criteria and conventions using model texts
	c. (Guided) practice	Apply discussed criteria to authentic tasks in various genres with clear communicative goals and intended audience	Provide help when needed through scaffolding and product feedback
		After writing: Give peer feedback and assess own text according previously discussed criteria	
3. Self-regulation skills	a. Observational learning	Observe/discuss/compare model(s), (teacher or peer) setting goals and monitoring progress in relation to goals during the writing process, observe/discuss/compare effect of self-regulation on the written product	Model self-regulation during writing, setting a goal for writing and monitoring progress towards this goal
	b. Explicit instruction	Listen actively, take notes	Explain the differences between various communicative goals, explain the importance of setting communicative goals for writing in advance, and show when and how during the writing process progress towards the communicative goal can best be monitored
	c. (Guided) practice	Set communicative goal before writing, monitor progress towards this goal during writing, regulate own writing process and adapt if necessary, evaluate written product in relation to communicative goal, revise if necessary.	Provide help when needed through scaffolding, and self-regulation feedback

Table 4
Tekster's General Lesson Format

Lesson phase	Learning and teaching activities
1	Goal of the lesson is explicitly stated (**3b**)
2	Plenary introduction in which specific characteristics of text type are addressed through modeling (**2a**), comparing model texts (**2a**), or explicit teacher instruction (**2b**)
3	Introduction of authentic writing assignment in which communicative goal and intended audience are explicated (**3b**)
4	Acronym for the strategy is explicitly named (**1b**)
5	Content is generated in keywords (first step of the strategy; gradual release of responsibility from **1a** to **1c**, **3a** to **3c**)
6	Content is generated in keywords (second step of the strategy; gradual release of responsibility from **1a** to **1c**, **3a** to **3c**)
7	Text is written using organized content (third step of the strategy; **1c**, **2c**, **3c**)
8[a]	Students' texts are read (fourth step of the strategy; **2a**)
9[a]	Students' texts are evaluated by answering evaluative questions and/or giving feedback (fifth step of the strategy; **2a**)
10[b]	Students' texts are revised on the basis of the received feedback (sixth step of the strategy; **3c**)

Note. Bold numbers refer to focus and mode of instruction as shown in Table 3.
[a] Only for Grades 5 and 6. [b] Only for Grade 6.

study was also provided. The second section of the teacher's manual contained the 16 lesson plans teachers were expected to implement during the 8-week intervention period. A DVD with peer modeling video clips and examples of teacher modeling was also provided.

The Tekster teacher's manual served as a guide during the training session. First, teachers learned about the program's theoretical framework, goals, and general structure. Then, they focused on specific instructional practices and effective lesson implementation. For instance, one characteristic of effective teacher feedback about writing is providing students with individualized comments, based on their areas of strength and need (Parr & Timperley, 2010). Therefore, during the training session, teachers learned about and collaboratively practiced the underlying skills needed to provide this type of feedback (i.e., accurately assessing the quality of students' texts and adapting comments accordingly). At the end of the training session, the authors stressed that it was very important for each teacher to carefully read the entire teacher's manual and watch the full DVD before implementing Tekster.

Intervention Fidelity

Several fidelity measures were used to determine whether teachers implemented Tekster as intended. Specifically, fidelity was operationalized three ways: number of lessons taught by each teacher, number of lessons completed by each student, and teachers' adherence to the lesson plans included in the teacher's manual. The strategies used to collect fidelity data included reviewing teachers' log books, reviewing students' workbooks, and observing classroom instruction.

Teacher logbooks. Each teacher was asked to maintain a logbook during Tekster implementation to document the number of lessons completed and the duration of each. After the intervention period, 75% of the logbooks were returned. Analysis of those data indicated teachers taught an average of 10 (out of the intended 16) Tekster lessons. The average number of minutes required to complete a lesson was 43, with a range of 29 to 58.

Student workbooks. We collected and reviewed students' workbooks after the intervention period to determine the number of lessons each student completed. A lesson was considered complete if a student's workbook contained a text corresponding with the practice writing task for that lesson. Analysis of these data revealed considerable variability in the number of lessons students completed. On average, students completed 10 lessons ($SD = 4$); however, 8% of students completed less than four lessons and 53% of students completed at least 10 lessons.

Classroom observations. Observations were conducted in two thirds of the classrooms (selected at random) in Group 1 and Group 2. Each observation lasted the full length of the lesson and took place approximately half-way through the intervention period. The observations for each group occurred over a 2-week period; thus, there was variation in the particular lessons observed. Ten trained undergraduate students served as observers in this study. Because each classroom was observed by only one person, the reliability of the observational data was not able to be assessed.

Our observation instrument was based on the work of Hintze, Volpe, and Shapiro (2002) and designed to gather two types of data: general adherence to the lesson plan and frequency of using two key instructional practices—teacher modeling and the writing strategy. To assess whether a lesson plan was being implemented as intended, every 20 s observers tallied whether a teacher was on task (i.e., executing the actions specified in the lesson plan for that phase of instruction) or off task (i.e., doing something unrelated to writing instruction). Each on task tally was further categorized as plenary (i.e., involving the whole class) or individualized (i.e., involving individual students or a small group of students). To measure the frequency of teacher modeling and strategy use, observers recorded the number of times a teacher modeled something for the class and the number of the times a teacher referred to the writing strategy acronym or steps.

Analysis of the observational data indicated teachers adhered closely to what was specified in the Tekster lesson plans. On average, teachers were on task 92% of the observed instructional time and their actions were consistent with the general framework and key elements of Tekster. For example, the division between plenary and individualized instruction was relatively equal (on average, 54% and 46%, respectively), as intended. Teacher modeling and use of the writing strategy were also evident (on average, 1.3 and 1.4 times per lesson, respectively).

Assessment of Writing Quality

Writing tasks. Because generalization of writing proficiency across genres is not warranted when scores are obtained with only one writing task (Bouwer, Béguin, Sanders, & Van den Bergh, 2015), we assessed students' writing skills at each measurement occasion using three different types of texts: descriptive (tasks a, d, g), narrative (tasks b, e, h) and persuasive (tasks c, f, i), as shown in Table 1. The three tasks for each genre were as similar as

possible and differed only in topic, not format. All nine writing tasks were developed by the authors for the purpose of this study and in conjunction with other experts in the field. To increase the likelihood students would produce texts of reasonable length, specific attention was given to ensure an appropriate level of difficulty and topical interest. For each task, students received a handout that included the written prompt, topically related image, and space for prewriting (if desired). A sample prompt for each text type is provided as Supplemental Appendix B in the online supplemental materials.

Administration of writing tasks. The writing tasks used as assessments in this study were administered by the participating teachers to students in their classroom during regular instructional time. Teachers were asked to administer the three writing tasks for each measurement occasion within one week, but not on the same day. Students completed each writing task independently and without a time constraint. Teachers were instructed not to provide students with any additional assistance while they completed the assessments.

Rating Writing Quality

We anonymized all student texts to reduce the likelihood characteristics such as gender or grade level would influence raters' judgments (Peterson, 2000). However, due to the scope of this study (1,420 students and nine writing tasks, resulting in approximately 12,780 written texts), it was not feasible to type students' handwritten work as a way to control for possible presentation effects (Graham, Harris, & Hebert, 2011). Global text quality was assessed using a continuous (interval) rating scale with five benchmarks (Blok & Hoeksma, 1984; Bouwer, Koster, & Van den Bergh, 2016). The midpoint on the scale was an average quality text, assigned an arbitrary score of 100. The other four benchmark texts were located one and two standard deviations above and below the midpoint and scored (in ascending order) as 70, 85, 115, and 130. A different benchmark scale was constructed for each text type. Supplemental Appendix C in the online supplemental materials contains a sample benchmark rating scale.

The rating scale benchmarks originated from a preliminary investigation of a randomly selected subsample drawn from all the texts (i.e., all three grade levels and genres) written during M1. Five experienced Grade 4–6 teachers rated the subsample holistically and their scores were averaged. Benchmarks were then selected based on two criteria: (a) the text was a good representation of the quality level ($-2\ SD$, $-1\ SD$, 0, $+1\ SD$, $+2\ SD$); and (b) the level of rater agreement about text quality was high.

The raters for the full assessment sample were also experienced Grade 4–6 teachers ($n = 47$). Raters were trained in advance how to use the benchmark scales and they were blind to experimental conditions. Each rater compared a student's text to the benchmarks and assigned a score, accordingly. Each text was rated by a jury of three people, using a design of overlapping rater teams. With this method, all the student texts were divided randomly into subsamples, equal to the number of raters. Each rater received three subsamples, based on a predetermined design. The overlap in subsamples allowed us to approximate the reliability of raters and juries (Van den Bergh & Eiting, 1989). The average reliability of jury ratings across tasks was high, $\rho = .89$, with the variation between tasks being $\rho = .86$–.91. The final quality score for each

text was determined by computing the mean of the three ratings. The raters' scores were normalized for each task using Blom's rank-based normalization formula (see Solomon & Sawilowsky, 2009) because they appeared to be negatively skewed (i.e., low quality texts tended to be scored more extremely).

Data Analyses

The data in this study were hierarchically organized; scores were cross-classified with students and tasks, and students were nested within classes. Therefore, the data were analyzed by applying different (cross-classified) multilevel models in which parameters were added systematically to the model. In such models, all students—including those with partly missing values—are taken into account.

The effectiveness of Tekster across groups and grade levels was tested with six models. Model 1 was the basic null model in which we only accounted for random error (S^2_e) and random effects of students (S^2_s), tasks (S^2_t), and classes (S^2_c). That is, writing scores were allowed to vary within and between students, between tasks (including systematic variation due to genre), and between classes. In Model 2, measurement occasion was added as a fixed effect to test whether average scores differed over time. Whether the variances within and between students, and between classes, differed between the three measurement occasions was tested in Model 3. In Model 4, group was added as a fixed effect to test whether average scores differed between the two groups. Model 5 tested the main effect of the intervention by estimating the interaction between group and measurement occasion. This model included the restriction that the effect of the intervention was the same in the two groups. Finally, this restriction was removed in Model 6 to test whether the intervention was equally effective in Group 1 and 2 which, in essence, provided a check on the internal validity of the experiment.

The maintenance effect of the intervention was tested by performing a specific contrast analysis of students in Group 1. In this analysis, students' posttest and delayed posttest scores were compared. To test whether the intervention was equally effective across grade levels, we applied two additional models. In the first model, grade was added as a fixed effect to test whether average scores differed between the three grades. In the second model, the interaction effect between the intervention (Measurement Occasion × Group) and grade level was added to test whether the intervention was equally effective across the three grades.

The role of gender on the effectiveness of the intervention was tested by two additional models. In the first model, gender was added as a fixed effect to test whether average scores differed between male and female students. In the second model, the interaction effect between the intervention (Measurement Occasion × Group) and gender was added to test whether the intervention was equally effective for males and females.

To test whether the intervention was equally effective for students with different levels of writing proficiency, we performed an aptitude treatment interaction analysis. For this analysis, the regression of students' pretest scores on their posttest outcomes was estimated per group.

Results

Effect of the Intervention

Results of the fit and comparison of the six models are shown in Table 5. There was a fixed effect of measurement occasion— Model 2 vs. Model 1, $\chi^2(2) = 279.61$, $p < .001$—which indicates average writing scores were not equal over time. Allowing the variances to differ between measurement occasions significantly improved the model—Model 3 vs. Model 2, $\chi^2(12) = 657.61$, $p < .001$; thus, for at least one level (students, tasks, classes, and/or random error), the variance was not homogeneous across measurement occasions. The main effect for group—Model 4 vs. Model 3, $\chi^2(1) = 1.32$, $p = .25$—was not statistically significant, indicating average scores were the same for students in Group 1 and 2.

There was a statistically significant effect for the intervention— Model 5 vs. Model 4, $\chi^2(1) = 24.98$, $p < .001$—as indicated by the interaction between group and measurement occasion. That is, differences in scores measured at two occasions (i.e., first and second or second and third) were not the same for students in the intervention and control conditions. The effect of the intervention on differences in scores appeared to be the same for students in Group 1 and 2—Model 6 vs. Model 5, $\chi^2(1) = 0.12$, p = .73. To verify the interaction effect between group and measurement occasion, we tested two additional contrasts. The interaction between group and the first two measurement occasions was statistically significant, $\chi^2(1) = 11.52$; $p < .001$; the difference in mean scores between measurement occasions was larger for Group 1. The interaction effect between group and the latter two measurement occasions was also statistically significant, $\chi^2(1) = 30.86$; $p < .001$; the difference in mean scores between the second and third measurement occasion was larger for Group 2.

Parameter estimates of Model 5 are summarized in Table 6 and a graphical display of the intervention effect is presented in Figure 1. The variance within and between students decreased over time, as did the variance between classrooms. The decrease in between-class variance means classes became more homogeneous over time. The reduction in within-student variance resulted from smaller interaction effects between students and tasks, indicating students' writing also became more homogeneous.

To estimate the magnitude of Tekster's effect, we compared the impact of the intervention to the total variance (Cohen's *d*). The overall ES (i.e., across all students, teachers, and tasks; based on the mean number of student-completed lessons) was 0.32. Because we discovered considerable variability in the number of lessons students completed ($M = 10$ and $SD = 4$, as reported above under

the Intervention Fidelity section), we included this variable as a fixed factor in the analyses. The results indicated a statistically significant, positive relationship between the number of student-completed lessons and the intervention effect, $\beta = 0.21$ ($SE = 0.09$, $p < .01$). On average, students who completed all 16 Tekster lessons had a gain score of 5.99, which translates to an ES of 0.40.

Maintenance. For students in Group 1, the impact of Tekster was measured immediately after the intervention period (M2) and again, 8 weeks later (M3). Results of the specific contrast analyses indicated the effect of the intervention was maintained over time. There was a statistically significant increase in students' scores between M1 and M3, $\chi^2(1) = 23.14$, $p < .001$, but the difference between M2 and M3 was not statistically significant, $\chi^2(1) = 2.06$, $p = .15$.

Grade level. The main effect for grade level was statistically significant, $\chi^2(2) = 54.40$, $p < .001$, meaning average scores differed for students in Grade 4, 5, and 6. The interaction between the intervention and grade level was also statistically significant, $\chi^2(2) = 14.21$, $p < .001$, indicating the impact of Tekster differed based on grade level. On average, Grade 4 students' scores increased by 4.86 points (ES = 0.34), Grade 5 students' scores increased by 5.00 points (ES = 0.35), and Grade 6 students' scores increased by 4.23 points (ES = 0.30). A graphical display of the intervention effect for each grade level is presented in Figure 2.

Gender. The main effect for gender was statistically significant, $\chi^2(1) = 319.70$, $p < .001$. On average, female students' scores exceeded male students' scores by 7.62 points. The effect of Tekster, however, was not gender dependent, as indicated by a nonsignificant improvement in the model when the interaction between group, measurement occasion, and gender was allowed, $\chi^2(1) = 0.10$, $p = .75$.

Writing proficiency. For students in Group 1, who participated in the intervention in the first 8 weeks, the regression coefficient of the scores of the first measurement occasion on the second measurement occasion equaled 0.60 ($SE = 0.02$). The regression coefficient for Group 2 in the same period equaled 0.59 ($SE = 0.03$), which was a nonsignificant difference ($t = 0.20$; $p = .42$). Hence, the results did not show an aptitude treatment interaction, indicating the effects of the intervention did not depend on students' writing proficiency.

Discussion

In this study, we tested the effectiveness of Tekster, a comprehensive, strategy-focused writing instruction program developed for Dutch students in Grades 4–6. Participating teachers imple-

Table 5
Fit and Comparison of Nested Models

Model	$N_{parameters}$	$-2LL$	Models	ΔX^2	Δdf	p
			Comparison			
1 null	5	88763.76				
2 + measurement occasion (fixed)	7	88484.15	2vs1	279.61	2	<.001
3 + measurement occasion (random)	19	87826.54	3vs2	657.61	12	<.001
4 + group	20	87825.22	4vs3	1.32	1	.25
5 + intervention	21	87800.24	5vs4	24.98	1	<.001
6 + Intervention × Group	22	87800.12	6vs5	.12	1	.73

Table 6

Students' Average Writing Scores and Variances on Pre- and Posttest Measures

Measure	Measurement occasion		
	1	2	3
Fixed part			
Group 1	95.63 (1.38)	100.36 (1.36)	99.36 (1.34)
Group 2	98.54 (1.41)	98.78 (1.33)	103.51 (1.28)
Random part			
$S^2_{classes}$	53.92 (11.31)	49.79 (10.44)	43.73 (9.40)
S^2_{tasks}	9.20 (1.42)	9.20 (1.42)	9.20 (1.42)
$S^2_{students}$	59.99 (4.33)	54.98 (3.65)	54.31 (3.65)
S^2_{error}	128.48 (3.68)	99.11 (2.83)	92.98 (2.77)

Note. Standard errors are included in parentheses.

mented the intervention in their general education classrooms for a period of 8 weeks. Results indicated Tekster produced statistically significant improvements in the quality of students' texts. Students' individual writing quality did not only increase, but also became more consistent over time. The switching replication design allowed us to replicate the effect of the intervention within this study, as our findings demonstrate that the intervention was equally effective in both groups. Moreover, we found that students in Group 1 still wrote qualitatively better texts at the delayed posttest measure than at the pretest measure, indicating that the effect of the intervention was maintained after two months. Although there was a significant improvement of students' writing scores in all grades, the effect of the intervention was slightly smaller in Grade 6 than in Grades 4 and 5. Furthermore, results show that girls outperformed boys on all measurement occasions, but that the effect of the intervention was the same. Lastly, results of an aptitude treatment analysis showed that the effect of the intervention did not depend on students' writing proficiency.

Although the intervention was effective, the ES of the intervention on students' writing was moderate, 0.32. However, this ES is based on the average of completed lessons (which was 10) and is
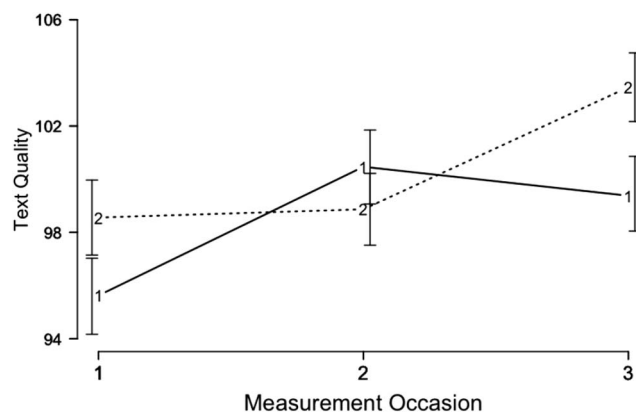

Figure 1. The effect of Tekster, averaged across all three grade levels. Error bars indicate 95% confidence intervals for the means. Solid lines represent Group 1, which received the intervention between first and second measurement occasion. Dashed lines represent Group 2, which received the intervention between second and third measurement occasion.
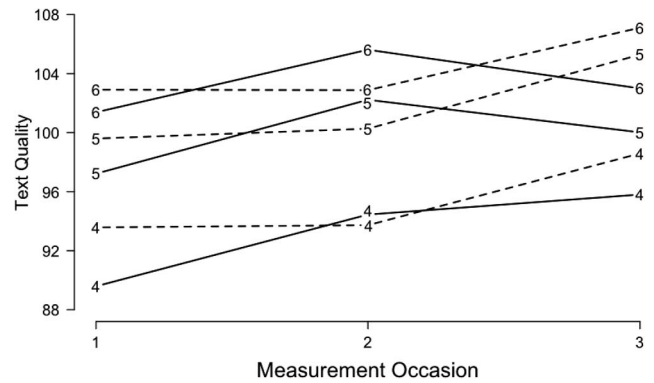

Figure 2. The effect of Tekster, by grade level. Solid lines represent Group 1, which received the intervention between first and second measurement occasion. Dashed lines represent Group 2, which received the intervention between second and third measurement occasion. Grade level is designated by the number (4, 5, or 6) shown with each line.

therefore a conservative estimate of the actual effect. Results showed the ES increased from 0.32 to 0.40 for students who completed all 16 lessons. Hence, students will make more progress if they complete the whole program. This can be achieved more easily when the implementation of Tekster is spread out over a longer period of time (e.g., one lesson a week), and/or if the program contained more lessons. Further research is needed to gain more insight in this aspect.

The effect of the intervention can also be interpreted in a more intuitive way by comparing it to the general improvement in writing skills of students between Grade 4 to 6 (Lipsey et al., 2012). Working with Tekster for 2 months resulted in an average gain in writing quality of 4.73 points. The average improvement in text quality scores between grades was 8.07 points, which means that students' writing improved by more than half a grade level.

Although Tekster was generally effective in improving students' writing performance, results showed that students' writing quality in Grade 4 and 5 improved slightly more than the writing quality of sixth grade students. An explanation for this can be that, even though the general approach is the same across grades, the acronyms differ slightly. Grade 6 is the only grade in which students are explicitly instructed to evaluate and revise. Research has shown that revising is difficult for students (Fitzgerald, 1987). To be able to revise, students must be aware of the goals and audience of their texts. In addition, they have to be able to critically read and evaluate their texts, and they have to know how they can fix problems, both on local and textual levels. Ideally, students start working with Tekster in Grade 4, when the focus is on learning and applying prewriting strategies, and gradually move on to Grade 6, when the focus shifts to revising. As this experiment was a cohort study, sixth graders lacked the basics that were the focus of instruction in Grade 4 and 5. We have addressed this issue by creating overlap in the topics that are covered in the different grades, but it might be that learning this overall approach at once was more complicated for Grade 6 students than the simpler versions of the acronym that were used in Grade 4 and 5. A longitudinal study would provide more insight in this matter.

A longitudinal study of Tekster would also shed more light on the learning trajectory of students across grades. The Dutch In-

spectorate of Education (Henkens, 2010) reported that at present students hardly progress in their writing from Grade 4 to 6. As we have developed a systematic approach for the teaching of writing in the upper primary grades, we would expect a more continuous development of students' writing performance across the grades as a result.

Generalizability of the Results

In comparison to similar strategy-focused intervention studies aimed at Grade 4 to 6 in a general educational setting, the ES of this study (0.32) is notably smaller (cf. Graham et al., 2012; Koster et al., 2015, average ES 1.02 and 0.96, respectively). However, in contrast to most other intervention studies, Tekster was tested on a very large scale involving 1,420 students from 60 classes from 27 schools. Moreover, whereas most intervention studies used only one task as an indication of the effectiveness of their writing program, we tested students' overall writing proficiency with nine writing tasks in three genres: narrative, persuasive, and descriptive. Effects are therefore not only generalizable across students, but also across teachers and tasks. If we were to ignore the variance component related to tasks and classes, the ES of our intervention would increase to 0.63 and to 0.80 if the full program would have been completed, which is more in line with the effects reported in other intervention studies.

Maintenance Effects

Our results show that students' writing quality is still significantly above pretest level two months after the end of the program, which suggests that the intervention induced a lasting change in students' writing. We also see that students' writing scores did not continue to gain after the end of the intervention period. This is a mere illustration of Henkens's observation (Henkens, 2010) that the regular writing lessons in the average Dutch classroom do not lead to any significant improvement in students' writing. This is demonstrated in the present study by the fact that students in the control group (i.e., Group 2 between the first and the second measurement occasion) did not show any gains in writing quality.

It should be noted, however, that conclusions about the maintenance effect of the intervention are true only under the assumption that tasks were equally difficult and the effect of the intervention (i.e., interaction between condition and time) was the same for students in both conditions. Naturally, we tried to keep the writing tasks as similar as possible over the three measurement occasions, using the same rating procedure in which raters used the same benchmark scale for equal tasks across occasions, and calculating average scores based on three writing tasks per occasion. Nonetheless, we cannot entirely rule out the possibility that differences or similarities between scores over time (within conditions) are due to coincidence.

Effectiveness of Tekster for Different Types of Students

Results did not show an aptitude treatment interaction, indicating that all students, less proficient as well as proficient writers throughout Grade 4 to 6, benefited from the program to the same extent. This suggests that the program addressed the needs of all students, which is promising, given that in a general education classroom students differ considerably in their needs and abilities (Harris et al., 2012). The effectiveness of the program for different types of students can be explained in at least three ways. First, Tekster aimed to reduce cognitive overload during writing by providing students with skills and knowledge to regulate their writing process. Second, the program addressed the double challenge of writing and learning to write at the same time. Third, through Tekster's multifaceted approach, all students, weak as well as proficient writers, were provided with ample learning opportunities, for example by including coping as well as mastery peer modeling (Braaksma, 2002). That Tekster enhances the performance of all students is promising for whole classroom use, as a typical upper elementary classroom will contain students of various abilities.

Tekster's Effective Components

It should be noted that, although the program as a whole improved students' writing performance, we cannot make claims about the effectiveness of its individual components. We simply do not know which component is the most powerful ingredient of our approach. What we do know from previous research is that the combination of strategy-focused instruction and observational learning is highly effective in improving students' writing performance (Fidalgo et al., 2015). Fidalgo and colleagues assessed the effectiveness of four different instructional components of a strategy-focused writing training: modeling and reflection, direct instruction, peer feedback, and individual practice for sixth grade students, by manipulating the instructional sequence. Their results indicated that all positive effects are predominantly related to the modeling and reflection component. The way our study was designed does not allow for any conclusions regarding the effect of the observational learning component, but based on Fidalgo et al.'s (2015) findings, we suspect that, especially in combination with strategy-focused instruction, modeling may have contributed substantially to the effectiveness of our program. However, additional research is needed to isolate the influence of each component.

Teachers' Implementation of Tekster

Tekster was implemented by fourth, fifth, and sixth grade teachers in their own general education classrooms. Teachers from a large variety of schools participated in the study. Although this contributed considerably to the ecological validity of this study, it increased differences between classes. Furthermore, differences between teachers can also be caused by differences in teaching experience, background, teaching styles and individual preferences (Hattie, 2009). Hence, it is important to verify how teachers actually implemented the program in their classrooms. In previous studies, researchers often controlled for the differences between teachers by implementing the intervention themselves (e.g., Gordon & Braun, 1986; Kellogg, 1988) or by training teachers or teacher assistants intensively to implement the intervention (e.g., Fidalgo et al., 2015; Graham et al., 2005). Whereas intensive training is possible in a relatively small-scale study of one or two classes, this is not a feasible option when an intervention is implemented on a large scale.

The differences between classes can partly be explained by differences in the number of taught lessons. On average, 10 lessons

were taught, but this number varied between classes, and we found that students' writing performance was positively related to the number of lessons taught. Furthermore, the results also showed that differences between teachers were reduced after the intervention, which suggests that teachers have adapted their instructional practice as a result of participation in the program. This seems to be confirmed by the fidelity measures, which revealed that teachers closely adhered to the lesson plans as indicated in the manual, and that they applied the key components of the intervention program, that is, modeling, the acronym, and the steps of the strategy.

It is promising that teachers were already capable of applying the key components of the program in their instruction after only a limited amount of training. However, the observational data do not allow for statements on the quality of the lessons, as they only provide information on what was done during the lessons. In further research, it is necessary to observe not only what teachers do in class, but also how they do this, for instance by videotaping and subsequently analyzing lessons to get a clearer picture of teachers' practices and whether and how they adapted the program to their own practice.

General Conclusion

To conclude, this study has shown that an overall approach in which several research-based instructional practices for teaching writing are combined is effective in improving elementary students' writing quality. This study is unique for the following reasons. First, through a switching replication design we were able to replicate the effect within one study, with the same results. Hence, the effects of the intervention do not seem to depend on characteristics of the sample. Together with the scale of the study, which included a large sample of Dutch schools, this allows us to make robust claims about the effectiveness of Tekster. Second, in this study we examined the impact of Tekster in a naturalistic setting, as the intervention was delivered in 60 general education classrooms by regular teachers, who were only trained for a short period of time. Third, students were taught a general strategy for writing, irrespective of genre, and the quality of their writing was measured with multiple writing tasks using multiple text types. It is therefore possible to generalize the results to overall writing proficiency in a general educational setting. All in all, this study demonstrates that a comprehensive writing program, such as Tekster, is a promising approach to improve elementary students' writing.

References

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

Bean, T. W., & Steenwyk, F. L. (1984). The effect of three forms of summarization instruction on sixth graders' summary writing and composition. *Journal of Reading Behavior, 16,* 297–306.

Bereiter, C., Burtis, P. J., & Scardamalia, M. (1988). Cognitive operations in constructing main points in written composition. *Journal of Memory and Language, 27,* 261–278. http://dx.doi.org/10.1016/0749-596X(88)90054-X

Berninger, V., Yates, C., Cartwright, A., Rutberg, J., Remy, E., & Abbott, R. (1992). Lower-level developmental skills in beginning writing. *Reading and Writing, 4,* 257–280. http://dx.doi.org/10.1007/BF01027151

Blok, H., & Hoeksma, J. B. (1984). *Opstellen geschaald: De constructie van beoordelingsschalen voor vijf schrijfopdrachten* [Scaling essays:

The construction of rating scales for five writing tasks]. Amsterdam, the Netherlands: Kohnstamm Institute.

Bodrova, E., & Leong, D. J. (1998). Scaffolding emergent writing in the zone of proximal development. *Literacy Teaching and Learning, 3,* 1–18.

Bouwer, R., Béguin, A., Sanders, T., & Van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing, 32,* 83–100. http://dx.doi.org/10.1177/0265532214542994

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

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

Braaksma, M. A. H., Rijlaarsdam, G., Van den Bergh, H., & Van Hout-Wolters, B. H. A. M. (2004). Observational learning and its effect on the orchestration of writing processes. *Cognition and Instruction, 22,* 1–36. http://dx.doi.org/10.1207/s1532690Xci2201_1

Brunstein, J. C., & Glaser, C. (2011). Testing a path-analytic mediation model of how self-regulated writing strategies improve fourth graders' composition skills: A randomized controlled trial. *Journal of Educational Psychology, 103,* 922–938. http://dx.doi.org/10.1037/a0024622

Central Office for Statistics. (2015, July 15). *(Speciaal) basisonderwijs; culturele minderheden, (achterstands)leerlingen* [(Special) primary education; cultural minority groups, (disadvantaged)students]. Retrieved from http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=37846SOL&D1=0&D2=a&D3=a&D4=a&HD=090218-1354&HDR=T,G2,G1&STB=G3

Couzijn, M. J. (1995). *Observation of writing and reading activities: Effects on learning and transfer* (Unpublished doctoral dissertation). University of Amsterdam, the Netherlands.

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

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

Crowhurst, M. (1990). Reading/writing relationships: An intervention study. *Canadian Journal of Education/Revue canadienne de l'éducation, 15,* 155–172. http://dx.doi.org/10.2307/1495373

Crowhurst, M. (1991). Interrelationships between reading and writing persuasive discourse. *Research in the Teaching of English, 25,* 314–338.

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

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

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

Ferretti, R. P., Lewis, W. E., & Andrews-Weckerly, S. (2009). Do goals affect the structure of students' argumentative writing strategies? *Journal of Educational Psychology, 101,* 577–589. http://dx.doi.org/10.1037/a0014702

Ferretti, R. P., MacArthur, C. A., & Dowdy, N. C. (2000). The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normally achieving peers. *Journal of Educational Psychology, 92,* 694–702. http://dx.doi.org/10.1037/0022-0663.92.4.694

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

Fitzgerald, J. (1987). Research on revision in writing. *Review of Educational Research, 57,* 481–506. http://dx.doi.org/10.3102/00346543057004481

Fitzgerald, J., & Markham, L. R. (1987). Teaching children about revision in writing. *Cognition and Instruction, 4,* 3–24. http://dx.doi.org/10.1207/s1532690xci0401_1

Fitzgerald, J., & Teasley, A. B. (1986). Effects of instruction in narrative structure on children's writing. *Journal of Educational Psychology, 78,* 424–432. http://dx.doi.org/10.1037/0022-0663.78.6.424

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

Gordon, C. J., & Braun, C. (1986). Mental processes in reading and writing: A critical look at self-reports as supportive data. *The Journal of Educational Research, 79,* 292–301. http://dx.doi.org/10.1080/00220671.1986.10885694

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

Graham, S., Harris, K., & Hebert, M. (2011). It is more than just the message: Analysis of presentation effects in scoring writing. *Focus on Exceptional Children, 44,* 1–12.

Graham, S., Harris, K., & Mason, L. (2005). Improving the writing performance, knowledge, and self-efficacy of struggling young writers: The effects of self-regulated strategy development. *Contemporary Educational Psychology, 30,* 207–241. http://dx.doi.org/10.1016/j.cedpsych.2004.08.001

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

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

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

Harris, K. R., Graham, S., Mason, L. H., & Saddler, B. (2002). Developing self-regulated writers. *Theory into Practice, 41,* 110–115. http://dx.doi.org/10.1207/s15430421tip4102_7

Harris, K. R., Lane, K. L., Graham, S., Driscoll, S. A., Sandmel, K., Brindle, M., & Schatschneider, C. (2012). Practice-based professional development for self-regulated strategies development in writing a randomized controlled study. *Journal of Teacher Education, 63,* 103–119. http://dx.doi.org/10.1177/0022487111429005

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, UK: Routledge.

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

Hillocks, G., Jr. (1984). What works in teaching composition: A meta-analysis of experimental treatment studies. *American Journal of Education, 93,* 133–170. http://dx.doi.org/10.1086/443789

Hintze, J. M., Volpe, R. J., & Shapiro, E. S. (2002). Best practices in systematic direct observation of student behavior. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 993–1006). Washington, DC: National Association of School Psychologists.

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

Inspectorate of Education. (2012). *De staat van het onderwijs. Onderwijsverslag 2010/2011* [The state of education. Educational report 2010/2011]. Retrieved from http://www.onderwijsinspectie.nl/binaries/content/assets/Onderwijsverslagen/2012/onderwijsverslag_2010_2011_printversie.pdf

Kellogg, R. T. (1988). Attentional overload and writing performance: Effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 355–365.

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

Koopman, P., Ledoux, G., Karssen, M., Van der Meijden, A., & Petit, R. (2015). *Vervolgmeting 1 kengetallen passend onderwijs* [Sequel measure 1 key ratios inclusive education] (Rapport 936, No. 20667). Amsterdam, the Netherlands: Kohnstamm Institute.

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

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

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

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

Koster, M., Tribushinina, E., De Jong, P. F., & Van den Bergh, H. (2015). Teaching children to write: A meta-analysis of writing intervention research. *Journal of Writing Research, 7,* 249–274. http://dx.doi.org/10.17239/jowr-2015.07.02.2

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

Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes, 50,* 212–247. http://dx.doi.org/10.1016/0749-5978(91)90021-K

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

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review, 8,* 299–325. http://dx.doi.org/10.1007/BF01464076

McCutchen, D. (2011). From novice to expert: Implications of language skills and writing-relevant knowledge for memory during the develop-

ment of writing skill. *Journal of Writing Research, 3,* 51–68. http://dx .doi.org/10.17239/jowr-2011.03.01.3

Midgette, E., Haria, P., & MacArthur, C. (2008). The effects of content and audience awareness goals for revision on the persuasive essays of fifth- and eighth-grade students. *Reading and Writing, 21,* 131–151. http://dx .doi.org/10.1007/s11145-007-9067-9

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97. http://dx.doi.org/10.1037/h0043158

Ministry of Education, Culture, and Science. (2015). *Basisonderwijs 2015– 2016* [Primary Education 2015–2016]. Den Haag, the Netherlands: Author.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. http://dx.doi.org/10 .1126/science.aac4716

O'Sullivan, J., & Pressley, M. (1984). Completeness of instruction and strategy transfer. *Journal of Experimental Child Psychology, 38,* 275– 288. http://dx.doi.org/10.1016/0022-0965(84)90126-7

Parr, J. M., & Timperley, H. S. (2010). Feedback to writing, assessment for teaching and learning and student progress. *Assessing Writing, 15,* 68–85. http://dx.doi.org/10.1016/j.asw.2010.05.004

Pearson, P. D., & Gallagher, G. (1983). The instruction of reading comprehension. *Contemporary Educational Psychology, 8,* 317–344. http:// dx.doi.org/10.1016/0361-476X(83)90019-X

Peterson, S. (2000). Grades four and eight students' and teachers' perceptions of girls' and boys' writing competencies. *Reading Horizons, 40,* 253–271.

Raedts, M., Rijlaarsdam, G., Van Waes, L., & Daems, F. (2007). Observational learning through video-based models: Impact on students' accuracy of self-efficacy beliefs, task knowledge and writing performances. In G. Rijlaarsdam, P. Boscolo, & S. Hidi (Eds.), *Studies in writing: Writing and motivation* (Vol. 19, pp. 219–238). Oxford, UK: Elsevier.

Raphael, T. E., & Kirschner, B. M. (1985). *The effects of instruction in compare/contrast text structure on sixth-grade students' reading comprehension and writing products* (Research Series No. 161). East Lansing, MI: Michigan State University, Institute for Research on Teaching.

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

Rijlaarsdam, G. (2005). Observerend leren: Een kernactiviteit in taalvaardigheidsonderwijs [Observational learning: A core activity in language education]. *Levende Talen Tijdschrift, 6,* 10–28.

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

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

Rijlaarsdam, G., Couzijn, M., Janssen, T., Braaksma, M., & Kieft, M. (2006). Writing experiment manuals in science education: The impact of writing, genre, and audience. *International Journal of Science Education, 28,* 203–233. http://dx.doi.org/10.1080/09500690500336932

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

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

Scardamalia, M., & Paris, P. (1985). The function of explicit discourse knowledge in the development of text representations and composing strategies. *Cognition and Instruction, 2,* 1–39. http://dx.doi.org/10.1207/ s1532690xci0201_1

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

Schriver, K. A. (1992). Teaching writers to anticipate reader's needs: A classroom pedagogy. *Written Communication, 9,* 179–208. http://dx.doi .org/10.1177/0741088392009002001

Schunk, D. H. (1987). Peer models and children's behavioral change. *Review of Educational Research, 57,* 149–174. http://dx.doi.org/10 .3102/00346543057002149

Schunk, D. H. (1990). Goal setting and self-efficacy during self-regulated learning. *Educational Psychologist, 25,* 71–86. http://dx.doi.org/10 .1207/s15326985ep2501_6

Schunk, D. H. (2012). Social cognitive theory. In D. Schunk (Ed.), *Learning theories: An educational perspective* (6th ed., pp. 117–162). Boston, MA: Pearson.

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

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

Troia, G. A., & Graham, S. (2002). The effectiveness of a highly explicit, teacher-directed strategy instruction routine: Changing the writing performance of students with learning disabilities. *Journal of Learning Disabilities, 35,* 290–305. http://dx.doi.org/10.1177/00222194020 350040101

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

Van Steendam, E., Rijlaarsdam, G., Van den Bergh, H., & Sercu, L. (2014). The mediating effect of instruction on pair composition in L2 revision and writing. *Instructional Science, 42,* 905–927. http://dx.doi .org/10.1007/s11251-014-9318-5

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Wong, B. Y., Hoskyn, M., Jai, D., Ellis, P., & Watson, K. (2008). The comparative efficacy of two approaches to teaching sixth graders opinion essay writing. *Contemporary Educational Psychology, 33,* 757–784. http://dx.doi.org/10.1016/j.cedpsych.2007.12.004

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 17,* 89–100. http://dx.doi.org/10.1111/j.1469-7610.1976 .tb00381.x

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

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