Teaching Children to Write: A Meta-analysis of Writing Intervention Research

Monica Koster[a], Elena Tribushinina[a], Peter F. de Jong[b] and Huub van den Bergh[a]

[a] Utrecht University | The Netherlands

[b] University of Amsterdam| The Netherlands

Abstract: It has been established that in the Netherlands, as in other countries, a majority of students do not attain the desired level of writing skills at the end of elementary school. Time devoted to writing is limited, and only a minority of schools succeed in effectively teaching writing. An improvement in the way writing is taught in elementary school is clearly required. In order to identify effective instructional practices we conducted a meta-analysis of writing intervention studies aimed at grade 4 to 6 in a regular school setting. Average effect sizes were calculated for ten intervention categories: strategy instruction, text structure instruction, pre-writing activities, peer assistance, grammar instruction, feedback, evaluation, process approach, goal setting, and revision. Five of these categories yielded statistically significant results. Pairwise comparison of these categories revealed that goal setting (*ES* = 2.03) is the most effective intervention to improve students' writing performance, followed by strategy instruction (*ES* = .96), text structure instruction (*ES* = .76), peer assistance (*ES* = .59), and feedback (*ES* = .88) respectively. Further research is needed to examine how these interventions can be implemented effectively in classrooms to improve elementary students' writing performance.

Keywords: writing, meta-analysis, intervention, composition, elementary school

journal of WRITING RESEARCH

Koster, M.,Tribushinina,E., De Jong, P.F., & Van den Bergh, B. (2015) .Teaching children to write: A meta-analysis of writing intervention research. *Journal of Writing Research, 7*(2), 249-274. http://dx.doi.org/10.17239/jowr-2015.07.02.2

Contact: Monica Koster, University of Utrecht, Utrecht Institute of Linguistics OTS, Trans 10, 3512 JK Utrecht| The Netherlands – m.p.koster@uu.nl.

1. Introduction

With the large-scale introduction of computers, tablets, and mobile phones Western society has rapidly become more literate over the last two decades. E-mail and text messages are replacing oral face-to-face and telephone communication, increasing people's need to be able to communicate adequately in writing. Individuals who do not sufficiently master the basic skills of writing will eventually encounter serious problems in participating fully in daily life. More than ever it is essential that children develop their writing competence at a young age, as writing skills play a crucial role in educational and occupational success (National Commission on Writing, 2003).

Despite the fact that composition skills are of vital importance for a successful academic and professional career, it was established that a majority of students in the Netherlands do not attain the desired level of writing skills (Henkens, 2010). A Dutch national assessment study demonstrated that at the end of elementary school (grade 6) most students were not capable of writing texts that sufficiently convey a single, simple message to a reader (Kühlemeier, Van Til, Feenstra, & Hemker, 2013). Further, this study showed that students hardly progress in their writing competencies from grade 4 to grade 6. A national writing assessment in the US yielded similar results: of all grade 8 students only one-third performed at or above proficient level (Salahu-Din, Persky, & Miller, 2008). This is a serious cause for concern, because weaker writers are at a disadvantage in their secondary school and college years, when writing becomes increasingly important as a tool for learning (e.g. Bangert-Drowns, Hurley, & Wilkinson, 2004).

Kühlemeier and colleagues (2013) found that in the Netherlands the time and attention devoted to writing education is limited at elementary school. At the same time the Dutch Inspectorate for Education (Henkens, 2010) concluded that only a minority of schools succeed in effectively teaching writing. Besides, during their own professional education, Dutch teachers do not receive adequate training in writing themselves, nor are they sufficiently prepared for teaching writing (Leeuw, 2006; Smits, 2009). Furthermore, it was established that language teaching materials (i.e. textbooks and teacher manuals) often do not provide sufficient directions for teachers to enable them to support their students' writing processes and to give proper feedback (Stoeldraijer, 2012). It can be concluded that an improvement in the way writing is taught at elementary school in the Netherlands is clearly required.

Above all, any improvement of the teaching of writing in elementary school must be based on interventions that have proven to be effective in enhancing the quality of students' written texts. The aim of this study was to identify effective instructional practices for teaching composition to students in the upper grades of elementary school. An increasing amount of research has been done on writing interventions, resulting in an accumulation of studies testing various instructional approaches. To gain insight into which instructional approaches are specifically effective for elementary

students in grade 4 to 6, we conducted a meta-analysis of experimental and quasi-experimental writing intervention studies aimed at students in the upper elementary grades. A meta-analysis is the designated method for this purpose, as the magnitude and the directions of effects of a large number of studies are reviewed in a systematical way.

In the field of writing research, a number of meta-analyses have already been conducted. Some of these analyses focused on a specific type of intervention: for instance, in a review of 39 studies conducted with students from grade 1 to 12, Graham (2006) found that strategy instruction significantly improved students' writing performance. In a meta-analysis on the process approach to writing, Graham and Sandmel (2011) analyzed 29 studies, involving students grade 1 to 12, and found that process writing instruction had a significant, but modest, positive effect on the quality of students' writing. Furthermore, three meta-analyses (Bangert-Drowns et al., 1993; Goldberg, Russell, & Cook, 2003; Morphy & Graham, 2012) investigated the effect of word processing on text quality in grade K to 12, and all found positive effect sizes for this type of treatment, especially for weaker writers.

So far, there have been three comprehensive meta-analyses of experimental and quasi-experimental writing intervention studies, investigating multiple treatments: firstly Hillocks (1984) investigating 60 studies ranging from elementary grades to the first year of college; secondly Graham and Perin (2007) examining 123 writing intervention studies with adolescents (grades 4-12); and, thirdly, Graham, McKeown, Kiuhara, and Harris (2012) analyzed 115 (quasi-) experimental studies involving elementary students, grade 1 to 6. All three analyses used slightly different intervention categories, due to differences in the populations under investigation. Despite this, there was substantial overlap in results. All three meta-analyses consistently found grammar instruction to have a negative effect on text quality with effect sizes [*ES*] ranging from -.29 (Hillocks, 1984) to -.41 (Graham et al., 2012). Hillocks (1984) and Graham and Perin (2007) both found sentence combining (combining simple sentences into more complex ones), with an *ES* of .35 and .46 respectively; the study of models (study and imitation of model pieces of writing), with an *ES* of .22 and .17; and inquiry (present students with data and initiate activities designed to help students develop skills or strategies for dealing with the data in order to write about it), with an *ES* of .56 and .28, to have a positive effect on students' writing performance. Further, Graham and Perin (2007), as well as Graham and colleagues (2012), found that the process approach to writing, (*ES* .09 and .40 respectively); strategy instruction, (*ES* = 1.03 and 1.02); prewriting activities, (*ES* of .42 and .54); peer assistance when writing, (*ES* = .70 and .89); setting product goals (*ES* of 1.00 and .76); and word processing, (*ES* = .56 and .47), all had a significant positive impact on text quality. In addition, in their elementary meta-analysis, Graham and colleagues (2012) identified seven other effective practices to improve the writing of elementary students writing: feedback (adult and peer), with respective effect sizes of .80 and .37, the use of creativity and imagery (*ES* = .70), text structure instruction (*ES* =

.59), teaching transcription skills (*ES* = .55), assessing writing (*ES* = .42), comprehensive writing programs (*ES* = .42), and extra writing time (*ES* = .30).

Rogers and Graham (2008) conducted a meta-analysis of 88 single subject design studies, and found, consistent with the results of the extensive meta-analyses of experimental and quasi-experimental studies, that strategy instruction, word processing, prewriting activities, goal setting, and sentence construction were effective in improving students' writing performance. Additionally, Rogers and Graham (2008) found that reinforcing writing productivity, teaching strategies for editing text, and teaching strategies for constructing paragraphs were effective for both typical and struggling writers. In contrast to other findings, Rogers and Graham (2008) found a positive effect for the teaching of grammar. As a possible explanation for this divergent finding Rogers and Graham suggested that weaker writers, opposed to typical writers, may have profited from specific grammar instruction, or that the teaching method (teacher modeling) may have contributed positively to the effect of grammar instruction.

The meta-analysis that we conducted can be regarded as a refinement of the previously conducted meta-analyses of writing instruction, as we specifically focused on effective instructional practices for beginning writers (grade 4-6) in a regular educational setting. All previous meta-analyses investigating multiple treatments included a broad range of students: all elementary grades (Graham et al., 2012), adolescents (Graham & Perin, 2007), or elementary to college students (Hillocks, 1984). We expected, however, that different types of treatment would be effective for different groups of students. It was our expectation that the effectivity of types of intervention would differ between elementary students, secondary students, and college students. Further, we even expected this to differ between lower and upper elementary students. Bourdin and Fayol (1994) have demonstrated that students until the fourth or fifth grade in general perform better orally than in writing, when producing narratives. Their study shows that young students, due to less automation, have to allocate their cognitive resources mostly to the low-level activities of writing, such as lexical access, sentence generating, and graphic execution, which interferes with the higher order skills, such as planning and content generation. Berninger, Yates, Cartwright, Rutberg, Remy, and Abbott (1992) have shown that in the early elementary grades students' writing performance is highly dependent on the degree to which the lower level skills that are conditional to writing are developed. In the upper grades of elementary school it is expected that these lower level skills have been automatized through maturation and practice, in such a way that students are able to focus on the composing process itself (Kress, 1994). It is anticipated that during this stage students will be more sensitive to instruction and practice in basic composition skills. Therefore we decided, unlike Graham and colleagues (2012), to exclude studies aimed at the lower levels of elementary school, and only include studies targeted at students grade 4 to 6.

Further, the prior analyses also included studies targeted at specific groups of students, for example struggling writers, learning disabled students, bilingual students, or high-achieving students. In our opinion, one should be cautious to generalize results

from studies targeted at such specific groups to the general population of all students in a regular school setting, as the instructional needs of these groups are bound to differ. For instance, bilingual students might need more grammatical and linguistic support, while the struggling writer might be helped more by instruction in mastering the basics of writing, whereas gifted students could need more challenging writing tasks and approaches. For this reason, we chose to include only studies focused at the full range of students in a regular classroom.

But, above all, none of the previous reviews went beyond summarizing effect sizes and statistically compared interventions to examine whether they differed significantly from each other in effectiveness. In that sense they could be considered statistical reviews more than that they provided answers on the level of differential effectiveness of specific interventions. With our analysis we expanded the previous meta-analyses by not only identifying effective interventions, but by also statistically determining their level of effectiveness by comparison.

Lastly, our meta-analysis can also be considered as an update of the previous body of meta-analytical research: a quarter of the studies we located were not included in prior meta-analyses.

In summary, the research question guiding this meta-analysis was: Which instructional practices effectively improve the writing performance of students in the upper elementary grades? To answer this question, we systematically reviewed 32 (quasi-) experimental writing intervention studies aimed at students grade 4-6. The findings from this meta-analysis have important implications for designers of teaching materials and teacher educators, on how the teaching of composition in upper elementary education can be improved.

2. Method

2.1 Inclusion criteria and search procedure

In order to be included in the meta-analysis, studies had to meet the following five criteria. First, the study had to involve students in the upper grades of elementary school (grade 4-6) in a regular school setting. Studies that were conducted in a special educational setting or only involving struggling writers were excluded from the analysis. Second, we only included experimental or quasi-experimental studies in which at least two instructional conditions were compared: an experimental condition and a control condition. This could either be a 'pure' control condition, in which no extra instruction was given, or a control condition in which an alternative treatment was provided. As a consequence, correlational and qualitative studies were excluded from this meta-analysis. Third, each study had to include a measure of text quality at posttest, as this provided the best indication of the impact of an intervention on students' writing performance. Scores for text quality are based on a reader's overall impression of the student's text, taking into account several factors, such as content, organization,

vocabulary, as well as style and tone. Other outcome measures, such as text length or students' motivation were reported only in some of the studies and could therefore not be included in the meta-analysis. Fourth, to be included in the analysis, studies had to provide the statistics necessary to compute a weighted effect size. Lastly, only studies presented in English were included in the meta-analysis.

The studies for this meta-analysis were located by searching the electronic databases of PsychINFO, ERIC and Google Scholar. For our study, we replicated the search procedure employed by Graham and colleagues (2012), using the keywords 'writing' or 'composition', combined with keywords indicating the type of 'intervention', such as: *assessment, collaborative learning, creativity, dictation, free writing, genre, goal setting, grammar, handwriting, imagery, inquiry, mechanics, models, motivation, peer collaboration, peers, planning, pre-writing, process approach, process writing, self-evaluation, self-monitoring, sentence combining, sentence construction, spelling, strategies, strategy instruction, summary, technology, word processing,* and *word processor*. Subsequently, we added the following keywords to locate potentially promising practices from recent research: *editing, feedback, intervention, modeling, observational learning, outline, outlining revising*, and *revision*. Further, references of previous meta-analyses, reviews, and obtained papers were examined for relevant studies.

Databases of theses, dissertations, and conference proceedings were searched for unpublished studies on the topic. Additionally, a cited reference search of previous reviews and meta-analyses was conducted in Web of Knowledge to identify relevant studies.

This search procedure yielded approximately 2000 results, of which titles and abstracts were closely examined. First, we removed all non-intervention studies, as well as all studies that were not aimed at grades 4-6. Next, we omitted all studies that were not experimental or quasi-experimental. Subsequently, we removed all studies that lacked a proper control condition, and finally we excluded all studies only investigating specific groups of students, such as, for example, struggling writers, learning disabled students, bilingual students, or high-achieving students. We located 37 studies that met all inclusion criteria. However, despite this, five studies did not provide the necessary statistics to calculate effect sizes. We have contacted the authors of these studies to obtain these statistics, but unfortunately received no reply. Regrettably, these studies had to be excluded. The procedure described above resulted in the location of 32 studies that were suitable to be included in our meta-analysis.

2.2 Coding procedure

To obtain an adequate description for each study included in the meta-analysis, we coded the following variables: grade, number of participants, description of experimental and control condition, publication type (journal/dissertation/report/conference presentation/paper), and the genre of posttest measure (expository/narrative/informative/persuasive). It should be noted that coding was restricted to posttest

measures, as we used these measures to calculate effect sizes. Further, we coded a number of variables of which we expected that they could account for heterogeneity in effect sizes between studies. For this purpose we coded: design of the study (random assignment or quasi-experimental), attrition (% of total sample), period (in days) and intensity (in minutes) of intervention, person providing instruction (researcher, teacher, teaching assistant), and random assignment of teachers to conditions. Due to considerable differences between the scoring procedures that were used, and differences in the interpretation of reliability of scoring, it was not possible to administer one overall reliability score per study. Therefore, we coded aspects of the studies of which is known that they are related to the reliability of writing quality scores: type of assessment of writing quality (holistic or analytical), number of writing tasks at posttest and number of raters assessing the quality of posttest measure (e.g. Rijlaarsdam et al., 2011). All studies were coded by the first author and a trained assistant, a random sample of ten studies (one third of the total sample) were coded by both coders, with 97% agreement.

2.3 Categorizing interventions

For the analysis, all studies were thoroughly examined and grouped according to their focus of intervention. Subsequently, studies with a comparable focus of intervention were grouped into categories, based on the categories used in previous meta-analyses (e.g. Graham & Perin, 2007; Graham et al., 2012; Hillocks, 1984). For our study, we maintained the following categories from these meta-analyses: strategy instruction, text structure instruction, peer assistance, process approach, feedback, grammar instruction, and prewriting activities. We decided to use 'goal setting' instead of 'product goals', because our sample also included a study involving the setting of process goals, as well as setting product goals. Two types of intervention in our sample did not fit into the categories that were used by previous reviews, therefore we added two new categories: evaluation and revision. This resulted in a total of ten intervention categories, which are summarized in Table 1. It should be noted that the intervention categories are not completely mutually exclusive, for instance, prewriting activities and revision also are components of the process approach and strategy instruction. We classified studies according to the main focus of instruction as described by the authors. For example, Bui, Schumaker, and Deshler (2006) characterize their intervention as a strategic writing program, in which they also apply the process approach to writing. As the emphasis of this intervention is on teaching students strategies for writing, it was decided to place this study in the strategy instruction category. Another example of a study of which the intervention has elements of more than one category is the study of Wong, Hoskyn, Jai, Ellis, and Watson (2008) which combines self-regulated strategy development with feedback. As the main intervention under investigation is strategy instruction, the study was placed in this category, rather than in the category feedback.

Table 1. Description of intervention categories

Category	Description
Strategy instruction	Explicit and systematical teaching of writing strategies
Text structure instruction	Explicit teaching of knowledge of the structure of texts
Peer assistance	Students engage in joined activities during (parts of) the writing process
Evaluation	Teaching students to evaluate their own work with specified criteria
Goal setting	Students are assigned specific product or process goals before writing
Feedback	Students receive comments from others on their writing
Grammar instruction	Explicit teaching of grammar and/or construction of sentences
Revision	Focus on revising draft versions
Prewriting activities	Students engage in activities before writing: generating content/planning
Process approach	Focus on writing process and subprocesses: planning-writing-revising

Strategy instruction involves the explicit teaching of strategies for planning, translating and revising. The majority of studies in this category uses the Self Regulatory Strategy Development (SRSD) model of Harris and Graham (1996), in which students are additionally taught self-regulation strategies to manage the writing process, as well as declarative and procedural knowledge about writing. *Text structure instruction* is the explicit teaching of the structure of a text in a specific genre, such as the organizational structure of a persuasive essay, the story constituents and interrelations of narrative texts, or a compare/contrast essay. Peer assistance involves studies where students have to collaborate during different stages (planning, formulating, or revising) of the writing process, or where some form of tutoring is applied. Evaluation involves teaching students how to reflect on and to assess their own work. Most studies in this category used the 6 (+1) Traits Writing Model, which was developed in the 1980's in the US (Northwest Regional Educational Library, 2013). The 6 (+1) Traits Writing Model asks students to assess their compositions on ideas, organization, voice, word choice, sentence fluency, conventions, and presentation by using reflective questions and rubrics. *Goal setting* involves assigning students goals for their writing before they begin: either a product goal (e.g. writing paragraphs), or a process goal (e.g. acquiring a learning strategy). *Feedback* involves studies in which students receive comments on (aspects of) their writing, either from the teacher or from a peer. *Grammar instruction* involves interventions that are aimed at the construction of correct sentences. *Revision* involves studies in which students receive instruction in improving draft versions of texts. *Prewriting activities* involve studies that focus on techniques for generating content and planning, such as brainstorming, or using graphic organizers. The process approach is a comprehensive intervention where students engage in cycles of planning,

formulating, and revising, and in writing for real audiences with real purposes. Instruction is often at individual level, tailored to the student's needs through mini-lessons, writing conferences, and teachable moments. Further, self-reflection and evaluation is stressed, to stimulate student's ownership of their written products. Students collaborate when writing, in a supportive and nonthreatening writing environment (Graham & Sandmel, 2011).

There were three studies in our sample, Arter, Spandle, Culham, and Pollard (1993), Saddler and Graham (2005), and Dejarnette (2008), comparing two intervention conditions. We calculated an effect size for both interventions and subsequently placed them in two intervention categories. Finally, a number of studies investigated multiple conditions, e.g. the study of Schunk and Swartz (1993) investigated the effectiveness of setting product goals, as well as the effectiveness of setting process goals. In these instances we calculated separate effect sizes for all conditions.

2.4 Calculation of effect sizes and statistical analysis

For each individual study included in the analysis an effect size was calculated for writing quality at posttest. If a holistic score was available, then this score was used to calculate the effect size. If writing quality was scored on separate aspects, such as organization, ideas, or word choice, separate effect sizes were calculated for each aspect and subsequently averaged into one single effect size. Means and standard deviations were used to obtain effect sizes. Effect sizes were calculated using Hedges' g (the standardized mean difference) by subtracting the mean performance of the control group at posttest from the mean performance of the treatment group at posttest, dividing by pooled standard deviation of the two groups. Hedges' g provides a slightly better estimate than Cohens d, especially for smaller sample sizes (Borenstein, Hedges, Higgins, & Rothstein, 2011).

For the meta-analysis a random effects model was used, as it was assumed that the true effect varied from study to study, due to differences in participants as well as differences in interventions and implementation of interventions. Rather than estimating one true effect size, a random effects model estimates the mean of a distribution of effects. This allows for generalization to populations beyond the included studies (Borenstein, et al., 2011). For each treatment category, an average effect size was calculated as well as the confidence interval and statistical significance of the obtained effect sizes. In this way the effect of various treatments could be compared. Additionally, a test of homogeneity was conducted, to determine whether variability in effect sizes was larger than expected based on sampling error alone. When the homogeneity test was statistically significant, a moderator analysis was conducted to determine whether the variability could be explained by identifiable factors, such as treatment duration, publication type, or grade.

2.5 Description of studies included in the meta-analysis

Table 2 contains a description of all studies included in the analysis and their effect sizes, grouped per intervention category. The intervention categories are ranked according to the amount of effect sizes they contain, starting with strategy instruction as the largest category (11 effect sizes). Within the categories, studies are arranged per grade, in alphabetical order. For each study the following information is given: reference, publication type, grade, number of participants, short description of intervention and control condition, genre of text written at posttest measure, and the effect size. As can be seen, there are seven categories containing four or less effect sizes. We acknowledge the fact that these sample sizes do not allow for firm conclusions. Nevertheless, for the sake of completeness, it was decided to retain these categories in the analysis, as this would at least provide an indication of the possible efficacy of these types of interventions. In total, we calculated 55 effect sizes from 32 studies, and divided them into 10 intervention categories.

Table 2. Description of included studies grouped per intervention category

Study	Publication type	Grade	N	Intervention	Genre	Effect size
				Strategy instruction (*k*=11)		
Brunstein & Glaser (2011)	J	4	115	Strategy instruction + self-regulation vs. strategy instruction	N	0.84
Glaser & Brunstein (2007) 1	J	4	72	Strategy instruction vs. didactic lessons in composition	N	0.48
Glaser & Brunstein (2007) 2	J	4	79	Strategy instruction + self-regulation vs. didactic lessons in composition	N	1.12
Mason et al. (2012) 1	J	4	47	Strategy instruction + self-regulation (TWA + PLANS) vs. no treatment	I	1.13
Bui et al. (2006)	J	5	99	Demand Writing Instruction Model vs. traditional writing instruction (+Prewriting activities)	n.s.	0.34
Barnes (2013) 1	D	5	178	WISE (Writing In School Every day) vs. no treatment	N,I,P	0.11
Barnes (2013) 2	D	5	189	WISE + professional development vs. no treatment	N,I,P	0.33
Mason et al. (2012)	J	5	48	Strategy instruction (TWA) vs. no treatment	N	0.81

Study	Publication type	Grade	N	Intervention	Genre	Effect size
Fidalgo et al. (2015)		6	41	Strategy instruction vs. normal curriculum	I	2.11
Torrance et al. (2007)	J	6	95	CSRI (Cognitive Self Regulation Instruction) vs. normal curriculum	I	3.57
Wong et al. (2008)	J	6	57	SRSD strategy instruction + CHAIR + adult feedback vs. CHAIR + constant training time	P	0.64
				Text structure instruction (*k*=9)		
Fitzgerald & Teasley (1986)	J	4	49	Instruction in story constituents and interrelations vs. dictionary use and word study	N	1.07
Gordon & Braun (1986)	J	5	54	Instruction in narrative structure vs. instruction in poetry writing	N	0.32
Bean & Steenwyk (1984) 1	J	6	41	Direct instruction rule-governed vs. advice to find main ideas	I	1.07
Bean & Steenwyk (1984) 2	J	6	39	GIST: direct instruction intuitive approach vs. advice to find main ideas	I	0.84
Crowhurst (1990)	J	6	46	Instruction model for persuasion + writing practice vs. group discussion activities	I	1.11
Crowhurst (1991) 1	J	6	50	Instruction model for persuasion + writing practice vs. reading novels and writing book reports	P	1.10
Crowhurst (1991) 2	J	6	50	Instruction model for persuasion + reading practices vs. reading novels and writing book reports	P	0.78
Crowhurst (1991) 3	J	6	50	One lesson persuasion vs. reading novels and writing book reports	P	0.34
Raphael & Kirschner (1985)	C	6	45	Instruction compare-contrast text structure vs. normal curriculum	I	0.26
				Peer assistance (*k*=9)		
Paquette (2008)	J	4	50	6 + 1 Traits model with cross-age tutoring vs. no extra instruction (+ Evaluation)	n.s.	1.27

Study	Publication type	Grade	N	Intervention	Genre	Effect size
Puma et al. (2007) 1	R	4	1249	Writing Wings (cooperative writing) vs. normal curriculum	N,I	0.07
Saddler & Graham (2005) 1	J	4	44	Sentence combining with peer assistance vs. grammar instruction	N	1.66
Puma et al. (2007) 2	R	5	347	Writing Wings (cooperative writing) vs. normal curriculum	N,I	0.03
Yarrow & Topping (2001) 1	J	5	14	Metacognitive strategy instruction with peer assistance (tutor) vs. metacognitive strategy instruction with no interaction	N	0.70
Yarrow & Topping (2001) 2	J	5	12	Metacognitive strategy instruction with peer assistance (tutee) vs. metacognitive strategy instruction with no interaction	N	0.52
Brakel Olson (1990) 2	J	6	41	Writing lessons + peer partner vs. writing lessons only	N	0.42
Hoogeveen (2013) 1	D	6	96	Specific genre knowledge + peer response vs. no extra instruction	N,E	1.11
Hoogeveen (2013) 2	D	6	93	General aspects of communicative writing + peer response vs. no extra instruction	N,E	0.30
				Evaluation (*k*=7)		
Collopy (2009)	J	4	100	6 Traits writing model vs. no extra instruction	N	0.31
Paquette (2008)	J	4	50	6 + 1 Traits model with cross-age tutoring vs. no extra instruction (+ Peer assistance)	n.s.	1.27
Tienken & Achilles (2003)	J	4	98	Skills and strategies to self-assess writing vs. no extra instruction	N	0.41
Ross et al. (1999)	J	4/5/6	296	Self-evaluation with rubrics + teacher feedback vs. normal curriculum development	N	0.74
Arter et al. (1994)1	C	5	132	6 Traits writing model vs. observation (normal curriculum) (+ Process approach)	E,N	0.20

Study	Publica-tion type	Grade	N	Intervention	Genre	Effect size
DeJarnette (2008)	D	5	131	6 + 1 Traits writing model vs. Writing workshop	N	0.73
Coe et al. (2011)	R	5	4134	6 Traits writing model vs. no extra instruction	E	0.01
				Goal setting (*k*=6)		
Schunk & Swartz (1993) 2	J	4	20	Process goal + progress feedback vs. general goal (+ Feedback)	E,N,I	3.03
Schunk & Swartz (1993) 2	J	4	20	Process goal vs. general goal	E,N,I	2,62
Schunk & Swartz (1993) 2	J	4	20	Product goal vs. general goal	E,N,I	1,05
Schunk & Swartz (1993) 1	J	5	30	Process goal + progress feedback vs. general goal (+ Feedback)	E,N,I	3.15
Schunk & Swartz (1993) 1	J	5	30	Process goal vs. general goal	E,N,I	2.66
Schunk & Swartz (1993) 1	J	5	30	Product goal vs. general goal	E,N,I	1.65
				Feedback (*k*=4)		
Schunk & Swartz (1993) 2	J	4	20	Process goal + progress feedback vs. general goal (+ Goal setting)	E,N,I	3.03
Schunk & Swartz (1993) 1	J	5	30	Process goal + progress feedback vs. general goal (+ Goal setting)	E,N,I	3.15
Holliway (2004) 1	J	5	55	Feedback + rating vs. one sentence	E	0.84
Holliway (2004) 1	J	5	48	Feedback + reading as the reader vs. one sentence feedback	E	0.69
				Grammar instruction (*k*=4)		
Saddler & Graham (2005) 1	J	4	44	Grammar instruction vs. sentence combining with peer assistance	N	-1.66
Gein (1991) 1	D	4	109	School grammar vs. direct writing	E,N	-0.05
Gein (1991) 2	D	4	110	Sentence construction vs. direct writing	E,N	0.06
Gein (1991) 3	D	4	111	School grammar vs. sentence construction	E,N	-0.11

Study	Publication type	Grade	N	Intervention	Genre	Effect size
				Revision (*k*=3)		
Brakel Olson (1990) 1	J	6	40	Revision instruction vs. no extra instruction	N	0.04
Brakel Olson (1990) 3	J	6	37	Revision instruction + peer partner vs. no extra instruction (+ Peer assistance)	N	0.85
Fitzgerald & Markham (1987)	J	6	30	Revision instruction vs. reading good literature	N	0.89
				Prewriting activities (*k*=3)		
Brodney et al. (1999) 1	J	5	51	Reading combined with prewriting vs. no extra instruction	E	0.93
Brodney et al. (1999) 3	J	5	49	Prewriting only vs. no extra instruction	E	0.17
Bui et al. (2006)	J	5	99	Demand Writing Instruction Model vs. traditional instruction (+ Strategy instruction)	n.s.	0.34
				Process approach (*k*=3)		
Arter et al. (1994)	C	5	132	Process approach vs. 6 Traits model (+ Evaluation)	E,N	-0.20
DeJarnette (2008) 2	D	5	131	Writing workshop vs. 6 + 1 Traits writing model (+Evaluation)	N	-0.73
Varble (1990)	J	6	128	Whole language group vs. traditional language instruction	I	0.16

Note: For ***study***, numbers behind the references indicate that effect sizes were calculated for multiple conditions, or groups; these effect sizes are reported separately. For ***Publication*** type, J: Journal, D: Dissertation, R: Report, C: Conference presentation, P: Paper. For ***Genre***, N: Narrative, E: Expository, I: Informative, P: Persuasive, n.s.: not specified. When a study is included in another category as well, this is mentioned in parentheses.

3. Results

First, a random effects model was used to obtain an overall average effect size for all studies included in the meta-analysis. This overall effect size was g = .72, with a 95% confidence interval of [.49 - .94]. As effect sizes are highly dependent on study characteristics, additional analysis was needed to establish whether the various effect sizes together in the sample provide a proper estimate of the effect size in the

population. This can be determined by conducting a homogeneity test. This test indicates if the variability in effect sizes is larger than the expected variability based on sampling error alone. As the studies in our sample varied widely in focus and approach, we expected significant heterogeneity, which was confirmed by the homogeneity test: $Q = 511.51$, $df = 54$, $p < .001$. This indicated that a common effect size for the total sample of studies could not be assumed.

First, we investigated possible publication bias by conducting a moderator analysis with publication type as moderator on all studies in the meta-analysis. This analysis yielded no significant result ($p = .22$), indicating that the effect sizes of studies published in peer reviewed journals did not differ systematically in their effect sizes from studies from other publication types. The next step in our analysis was to examine the effectiveness of the various intervention categories, by including these 10 categories in our model as explanatory variables. The inclusion of the intervention categories significantly improved the model, according to a likelihood ratio test, with $X^2 = 19.69$, $df = 9$, $p < .001$. This means that differences in effect sizes were (at least partly) explained by the type of intervention.

Table 3. Summary of statistics for intervention categories

Intervention	N	Average	SE	95% Confidence		Heterogeneity	
				Lower	Upper	Q	I^2
Strategy instruction	11	0.96 ***	0.19	0.59	1.33	109.99***	94.30
Text structure	9	0.76 ***	0.21	0.34	1.18	11.91	33.87
Peer assistance	9	0.59 **	0.21	0.17	1.01	56.05***	89.83
Evaluation	7	0.43	0.23	-0.01	0.87	66.56***	87.57
Goal setting	6	2.03 ***	0.33	1.37	2.68	13.47*	62.61
Feedback	4	0.88 *	0.38	0.14	1.61	25.08***	91.08
Grammar	4	-0.37	0.30	-0.97	0.22	20.16***	91.84
Revision	3	0.58	0.38	-0.17	1.33	4.14	51.59
Prewriting activities	3	0.13	0.36	-0.58	0.85	3.91	48.57
Process approach	3	-0.25	0.34	-0.92	0.41	12.78**	84.58

Note: *** $p < .001$, ** $p < .01$, * $p < .05$

Table 3 gives the summary statistics for all intervention categories, presented in the same order as Table 2. These statistics include, per intervention category, the number of effect sizes, the average effect size and standard error, the 95% confidence interval, and the heterogeneity statistics Q (test statistic for heterogeneity) and I^2 (percentage of total heterogeneity/variability).

As can be seen in Table 3, we found two negative effects, for grammar instruction and the process approach. These interventions did not improve the quality of students'

writing. However, all other main effects were positive. Of these positive effects, five main effects significantly deviated from zero. These were, in order of effect size: goal setting, strategy instruction, feedback, text structure instruction, and peer assistance. Post-hoc analysis was conducted by a contrast analysis in which all interventions were compared pairwise. Results from these analyses showed that goal setting was by far the most effective intervention ($X^2 \geq 36.81$, $df = 1$, $p < .001$). However, as can be seen in Table 2, all effect sizes in the category goal setting were calculated from one study in which multiple conditions and grades were compared (Schunk & Swartz, 1993). This result should therefore be interpreted with caution. Goal setting was followed by strategy instruction ($X^2 \geq 26.06$, $df = 1$, $p < .001$), text structure instruction ($X^2 \geq 12.82$, $df = 1$, $p < .001$), and peer assistance ($X^2 \geq 7.64$, $df = 1$, $p = .006$) respectively. These three categories were all based on nine or more effect sizes from different studies. Feedback also proved to be an effective intervention, however, not more effective than prewriting activities.

The homogeneity test indicated that there was still a significant amount of residual heterogeneity in the sample ($QE = 283.18$, $df = 45$, $p < .001$). Therefore, we inspected the funnel plot (see Figure 1) to locate outliers that could be a potential source of heterogeneity. A funnel plot is a scatterplot of the intervention effect against a measure of study size. In the funnel plot in Figure 1 the residuals of the model with the intervention categories as explanatory variables were plotted against the standard error. The straight lines in Figure 1 define the region in which 95% of the studies was expected, in the absence of homogeneity. It can be seen that the studies were more or less symmetrically spread around the overall average effect size, and that most points were located in the region between the straight lines. This was an indication that there was no systematic heterogeneity in our sample. Two outliers (6.25% of the total sample) were located. The forest plot that we subsequently created (see Appendix), identified these outliers to be the studies of Torrance et al. (2007), and Saddler and Graham (2005). The effect size in the study of Torrance et al. (2007) was underestimated in the analysis whereas the observed effect size in the study of Saddler and Graham (2005) was smaller than expected, which meant that the effect size of the first study was larger, whereas in the latter study the effect size was smaller than in comparable studies (see also Figure 1). The analysis was repeated without these studies, but the outcome of this analysis did not significantly differ from the previous analysis ($X^2 = 3.61$, $df = 2$, $p = 0.16$). Hence, it was decided to maintain these studies, and to retain the previously estimated model for further analysis.

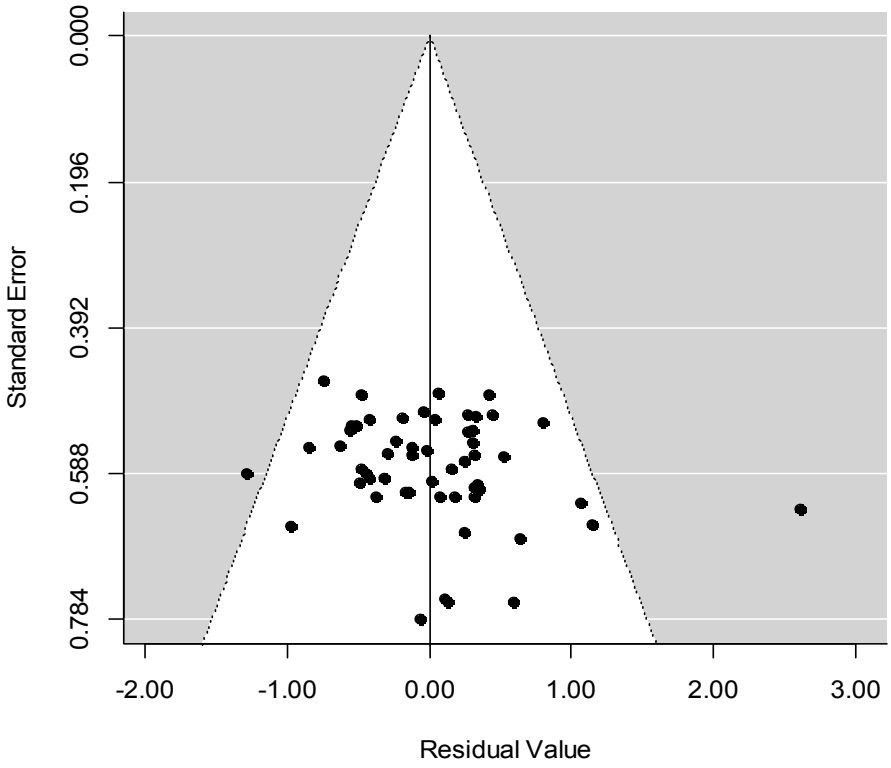

Figure 1: Funnel plot of final model.

Subsequently, a moderator analysis was conducted to examine whether the variability between studies could be attributed to one or more identifiable factors. We examined whether there were systematical differences in effect sizes between studies with a proper control condition and studies comparing different intervention conditions. In six intervention categories there were one or more studies without a no extra instruction control condition. Contrary to expectations, the inclusion of control condition as a moderating variable did not result in a significant reduction of residual heterogeneity (QE = 220.37, df = 37, p < .001), and for none of the intervention categories the parameter estimate for control condition was significant (p-values ranging from .29 to .90). Next, grade, duration of intervention, type of assessment of writing quality (holistic or analytical), number of writing tasks in posttest, and number of raters assessing the quality of posttest measure were considered as moderating factors. None of these factors significantly reduced the heterogeneity between studies in the total sample.

In the next step of the analysis, closer examination of the intervention categories separately revealed no significant heterogeneity in four categories: text structure instruction, process approach, revision, and prewriting (*p*-values ranging from .08 to .16). We further investigated the heterogeneity within the remaining intervention categories. This analysis was limited to categories containing more than five effect sizes, i.e. strategy instruction, peer assistance, evaluation and goal setting, as in the smaller categories the heterogeneity can largely be attributed to differences between individual studies. In the larger categories systematic factors may have caused heterogeneity, and this was examined by performing a moderator analysis on the separate categories with grade, duration of intervention, type of assessment of writing quality (holistic or analytical), number of writing tasks in posttest, and number of raters assessing the quality of posttest measure as potential moderators.

In strategy instruction, grade appeared to be a significant moderator: effect sizes were systematically larger in grade 6 (2.19) than in either grade 4 or 5 (0.59). Further, we found that effect sizes in this category were smaller (-0.86) for studies in which text quality at posttest was assessed analytically compared to studies in which holistic assessment was applied. In the category evaluation, genre of posttest was a significant moderator: effect sizes were smaller (-0.11) for expository texts. In the category peer assistance, heterogeneity could largely be attributed to one large study (Puma et al., 2007) with a relatively low effect size. In goal setting, heterogeneity could be attributed to differences between conditions.

However, from the 95% confidence interval statistics reported in Table 3 can be concluded that, despite significant heterogeneity within the categories of interventions that significantly improve writing proficiency, the effects in these categories were still largely positive, even at the lower bound of the confidence interval.

4. Discussion

4.1 Effective interventions to improve elementary students' writing

It has been established that, in the Netherlands, the way writing is taught in elementary school needs to be improved. The aim of this meta-analysis was to identify evidence-based effective instructional practices for teaching writing to students in grade 4 to 6. To determine this, we calculated average effect sizes for 10 types of interventions. The results show that the most effective interventions to improve students' writing are, in order of effect sizes: goal setting, strategy instruction, text structure instruction, feedback, and peer assistance. Post hoc analysis demonstrates that goal setting is the most effective intervention, followed by strategy instruction, text structure instruction, peer assistance, and feedback. This is in line with the findings of recent previous reviews (Graham & Perin, 2007; Graham et al., 2012), even though we limited our analysis to students in grade 4 to 6 in a regular educational setting. However, our findings are corroborated by statistical analysis.

The results of our analysis show that goal setting was by far the most effective intervention. However, as stated before, it should be noted that all effect sizes in this category come from one (twenty year old) study (Schunk & Swartz, 1993), comparing multiple conditions and multiple grades. Thus, these results only allow for tentative conclusions. Support for the positive effect of setting product goals can be found in previous meta-analyses (Graham & Perin, 2007; Graham et al., 2012) albeit for (partly) different populations of students (special needs learners, struggling writers, and slightly older students). This indicates that setting goals could help to improve students' writing.

Strategy instruction is the next effective intervention. Strategy instruction is the largest intervention category in our analysis, which allows for robust conclusions. Of all types of intervention, strategy instruction is by far the most investigated. It should be noted that that the majority of studies in this category examined the self-regulated strategy development (SRSD) approach of Harris and Graham (1996) to strategy instruction or a variation thereof. The SRSD approach seems to have developed into the 'standard' in strategy instruction, which is hardly surprising as studies examining SRSD invariably yield large effect sizes. Previous meta-analyses (Graham, 2006; Graham & Perin, 2007; Graham et al., 2012) also found SRSD to be a highly effective intervention for all types of learners (struggling writers, learning disabled, average, gifted) in a wide range of grades (grade 2 to 10). A subsequent moderator analysis, which we performed in all categories containing more than five effect sizes from different studies, shows that, in our sample, in grade 6 the (average) effect of strategy instruction appears to be much higher than in either grade 4 or 5. A possible explanation for this finding may be that in grade 6 students' lower level skills have been developed to such an extent that they profit the most from the explicit teaching of writing strategies. Further, we find that effect sizes in this category are smaller in studies where text quality is assessed analytically, compared to studies in which holistic assessment is used. In analytical assessment scoring rubrics are used: a set of criteria and standards that are linked to the learning objectives of the task at hand. Therefore, analytical assessments are more task-specific than a holistic assessment, which makes them harder to generalize to writing proficiency (Schoonen, 2005; Rijlaarsdam et al., 2011). As all different aspects of a text are evaluated separately, and subsequently combined into one final total score, analytical scores tend to be lower than holistic scores (Schoonen, 2005).

The next effective intervention category is text structure instruction. This category is a homogeneous sample of studies. The studies in this category investigate the effect of explicit teaching of (elements of) text structure, in different types of texts: narrative, persuasive, and compare-contrast texts. In all studies in this category the explicit teaching of text structure leads to a significant improvement of students' writing performance.

Text structure instruction is followed by peer assistance. Peer assistance is a diverse category: collaboration between students is applied in different phases of the writing process, with diverse types of interventions. As can be seen in Table 2, the effect of peer assistance depends on how it is applied, and on the focus of the intervention.

Studies with mainly cooperative writing (e.g. Puma et al., 2007) have smaller effects than studies combining peer assistance with more targeted types of interventions, such as the teaching of specific genre knowledge (Hoogeveen, 2013) or sentence combining (Saddler & Graham, 2005). Peer tutoring is also an effective practice to improve students' writing, as is shown by the study of Yarrow and Topping (2001). This study further shows that the writing scores of tutors improved more than those of the tutees. An explanation for this result may be that students learn more from explaining the material to others: you can only adequately explain something if you understand it yourself.

With only four effect sizes from two studies, feedback is one of the smaller intervention categories. Although seemingly effective, more research is needed to allow for more robust conclusions, as feedback can take many forms (e.g. peer feedback vs. teacher feedback) and can be applied in different ways (e.g. product-focused vs. process-focused). Further research should examine how and in what form feedback can be applied in teaching writing to improve students' writing performance.

Grammar instruction and the process approach to writing yield negative average effect sizes. The negative effect for grammar instruction confirms the findings in previous meta-analyses (Graham & Perin, 2007; Graham et al., 2012; Hillocks, 1984). Apparently, attention for the construction of correct sentences does not lead to improvement in text quality. This may be due to lack of transfer effects: when grammar is taught in isolation, and not in a 'real' writing context, it may not be clear to students how to apply what they learned when writing a text.

The negative effect for process approach may be explained by several factors. First, it is a small, but nevertheless homogeneous, intervention category of only three studies. In two out of these three studies, process approach is the control condition, thus compared with another (in this case: more effective) intervention type. We suspected that this could have resulted in lower effect sizes than when process approach would have been compared to a 'pure' control group. However, our suspicions were not confirmed by subsequent analysis with type of control condition as a moderator. There are several possible explanations for this result: the most straightforward one is that there are indeed no differences, but it can also be that our sample is too small and therefore lacking power to reveal systematic differences. However, it can also be that the process approach is too comprehensive for beginning writers: working on too many aspects at the same time. Beginning writers may profit more from a targeted intervention, such as text structure or strategy instruction. It must be noted that Graham and Perin (2007) found a (small) positive effect, for the process approach in their meta-analysis for adolescent students. This might indicate that the process approach is an effective approach for teaching writing to more experienced writers, but that this approach is less suitable for beginning writers.

4.2 Limitations of the study

We recognize the fact that some categories were small (≤ 4 effect sizes) and therefore only allow for tentative conclusions on the overall effectivity of these intervention categories. Nevertheless, since we wanted to obtain as much information as possible from the available data, these categories were included in the analysis, to examine their potential effectiveness.

A complicating factor in interpreting the results of the analysis is the fact that there was a considerable heterogeneity between studies that could not fully be accounted for by identifiable factors. However, it should be noted that the heterogeneity is overestimated due to the amount of small studies in our sample. A large amount of small studies in a category results in considerably more heterogeneity between studies, whereas in larger studies there is more heterogeneity within the study, and less between studies. In our sample, we see that the heterogeneity in the smaller categories is often caused by differences between individual studies. For instance, variation between studies may have been caused by differences in operationalization, such as the materials that were used, and the instruction that was given. Further, the number and nature of assignments that students had to work on varied considerably: from one writing task in one genre, to several writing tasks in various genres. The period of intervention varied even more: from one day to one year. A complicating factor in the analysis is that key aspects, such as the control condition, the nature of the posttest, the exact period of intervention, and the exact time spent on the intervention were not always described explicitly, which made it troublesome to code for these variables. These aspects can contribute to heterogeneity, but they cannot be included in a meta-analysis in a meaningful way if they are not reported accurately.

4.3 Suggestions for further research

From our study it becomes clear that there is not much writing intervention research conducted for students in the upper grades of elementary school. We can conclude that more research in this area is clearly needed. Some of the intervention categories in our meta-analysis were too small to draw firm conclusions about their effectiveness. Especially in these categories more research should be conducted. Particularly the effectiveness of goal setting needs further investigation, as our results indicate that it might be very effective in improving writing. It would certainly be worthwhile to investigate whether the positive results of the study of Schunk and Swartz (1993) can be replicated in other studies. But also feedback and prewriting activities need to be studied more closely. Additionally it should be examined if a combination of highly effective interventions will lead to even better student outcomes. In other words: is it worthwhile to add one highly effective intervention to another highly effective intervention, or does this only lead to marginal improvement? Further, other types of interventions, and new approaches should be developed and tested.

Furthermore, of the studies in our sample, 34% employs a posttest-only design and 47% a pretest-posttest design in which the effect is measured directly at the end of an

intervention. However, to make substantiated claims about the effectiveness of an intervention, a delayed posttest should be included to measure retention. Often, the posttest closely resembles what is taught during the intervention, which may lead to an overestimation of the effects. A delayed posttest could provide more information on the long-term effects of interventions on students' writing. Therefore, to make any claims about the 'real' effectiveness of interventions, delayed posttest data are essential. Unfortunately, this is still not common practice in intervention research.

4.4 Recommendations for teaching

This meta-analysis provides some valuable clues as to what works in teaching writing. Certainly more in-depth research is needed into what specifically works and what not, but we were already able to identify promising interventions for successfully teaching writing to students in the upper grades of elementary education. On the basis of our results, we must conclude that, to successfully improve the quality of writing of beginning writers, the writing curriculum should include goal setting, strategy instruction, text structure instruction, feedback and peer interaction. Setting process goals, such as learning to apply a certain strategy, was highly effective. Strategy instruction was more effective when combined with teaching self-regulatory skills. Overall, we found that specific, targeted interventions, such as explicit instruction in applying strategies or how to structure a text were particularly effective for elementary students. What we still do not know, is what the ideal instructional program for teaching composition skills should look like: which materials should we use, how much students have to write, how much practice students need, how we support the students' writing process, how we give appropriate feedback, and so on. In that respect, this analysis provides only rough guidelines for teaching, not a ready to use panacea. To determine what really works, extensive testing in classrooms is still needed.

Acknowledgements

This research has been supported by the Netherlands Organization for Scientific Research (NWO), grant 411-11-859.

Note

As this information was available for all studies, we restricted ourselves to calculating effect sizes for posttest only. Pretest information was available for 66% of the studies in our sample, whereas 19% of the studies also included a delayed posttest measurement. Pretest measures were only coded to verify whether there were pre-intervention differences between conditions, as most studies had a quasi-experimental design.

References

References marked with * are studies included in the meta-analysis

*Arter, J. A., Spandel, V., Culham, R., & Pollard, J. (1994). *The Impact of Training Students To Be Self-Assessors of Writing.* New Orleans. Paper presented at AERA.

Bangert-Drowns, R.L., Hurley, M.M.,& Wilkinson, B. (2004). The effects of school-based Writing-to-Learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74*, 29-58. http://dx.doi.org/10.3102/00346543074001029

*Barnes, J. C. (2013). *The Effects of a Writing Intervention on Fifth-Grade Student Achievement* (Doctoral dissertation).

*Bean, T. W., & Steenwyk, F. L. (1984). The effect of three forms of summarization instruction on sixth graders' summary writing and comprehension. *Journal of Literacy Research*, *16*(4), 297-306. http://dx.doi.org/10.1080/10862968409547523

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

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. West Sussex: Wiley.

Bourdin, B., & Fayol, M. (1994). Is written language production more difficult than oral language production? A working memory approach. *International Journal of Psychology, 29*(5), 591-620. http://dx.doi.org/10.1080/00207599408248175

*Brakel Olson, V.L. (1990). The revising processes of sixth-grade writers with and without peer feedback. *The Journal of Educational Research, 84*(1), 22-29. http://dx.doi.org/10.1080/00220671.1990.10885987

*Brodney, B., Reeves, C., & Kazelskis, R. (1999). Selected prewriting treatments: Effects on expository compositions written by fifth-grade students. *The Journal of Experimental Education*, *68*(1), 5-20. http://dx.doi.org/10.1080/00220979909598491

*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*(4), 922-938. http://dx.doi.org/10.1037/a0024622

*Bui, Y. N., Schumaker, J. B., & Deshler, D. D. (2006). The Effects of a Strategic Writing Program for Students with and without Learning Disabilities in Inclusive Fifth-Grade Classes. *Learning Disabilities Research & Practice*, *21*(4), 244-260. http://dx.doi.org/10.1111/j.1540-5826.2006.00221.x

*Coe, M., Hanita, M., Nishioka, V., & Smiley, R. (2011). *An Investigation of the Impact of the 6+ 1 Trait Writing Model on Grade 5 Student Writing Achievement. Final Report. NCEE 2012-4010*. National Center for Education Evaluation and Regional Assistance.

*Collopy, R. M. (2008). Professional development and student growth in writing. *Journal of Research in Childhood Education*, *23*(2), 163-178. http://dx.doi.org/10.1080/02568540809594653

*Crowhurst, M. (1990). Reading/writing relationships: An intervention study. *Canadian Journal of Education/Revue canadienne de l'éducation*, *15*(2), 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*(3), 314-338.

*DeJarnette, N. K. (2008). *Effect of the 6+ 1 Trait Writing Model on Student Writing Achievement*. ProQuest.

*Fidalgo, R., Torrance, M., Rijlaarsdam, G., Bergh, H. van den, 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.

*Fitzgerald, J., & Teasley, A. B. (1986). Effects of instruction in narrative structure on children's writing. *Journal of educational psychology*, *78*(6), 424-432.http://dx.doi.org/10.1037/0022-0663.78.6.424

*Fitzgerald, J., & Markham, L. R. (1987). Teaching children about revision in writing. *Cognition and Instruction*, *4*(1), 3-24. http://dx.doi.org/10.1207/s1532690xci0401_1

*Gein, J. van de (1991). *The sense of sentences. A study into the effects of grammar instruction upon junior writing*. Doctoral dissertation. Utrecht: Utrecht University.

*Glaser, C., & Brunstein, J. C. (2007). Improving fourth-grade students' composition skills: Effects of strategy instruction and self-regulation procedures. *Journal of Educational Psychology*, *99*(2), 297-310. http://dx.doi.org/10.1037/0022-0663.99.2.297

Goldberg, A., Russell, M., & Cook, A. (2003). The effect of computers on student writing: A meta-analysis of studies from 1992 to 2002. *Journal of Technology, Learning, and Assessment, 2*(1).

*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*(5), 292-301. http://dx.doi.org/10.1080/00220671.1986.10885694

Graham, S. (2006). Strategy instruction and the teaching of writing. In C. A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (1st ed., pp. 187-207). New York: The Guilford Press.

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445-476. http://dx.doi.org/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. http://dx.doi.org/10.1080/00220671.2010.488703

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

Harris, K., & Graham, S. (1996). *Making the writing process work: Strategies for composition and self-regulation* (2nd ed.). Cambridge, MA: Brookline Books.

Henkens, L. (2010). *Het onderwijs in het schrijven van teksten* [*Education in text writing*] (pp. 1–58). Utrecht: Inspectie van het Onderwijs.

Hillocks, G. (1984). What works in teaching composition: A meta-analysis of experimental treatment studies. *American Journal of Education, 93*(1), 133-170. http://dx.doi.org/10.1086/443789

*Holliway, D. R. (2004). Through the eyes of my reader: A strategy for improving audience perspective in children's descriptive writing. *Journal of Research in Childhood Education*, *18*(4), 334-349. http://dx.doi.org/10.1080/02568540409595045

*Hoogeveen, M. (2013). *Writing with Peer Response using Genre Knowledge*. Doctoral Dissertation. SLO: Enschede.

*Knudson, R. E. (1991). Effects of instructional strategies, grade, and sex on students' persuasive writing. *The Journal of Experimental Educational*, 141-152. http://dx.doi.org/10.1080/00220973.1991.10806557

Kress, G. (1994). *Learning to write*. London: Routledge.

Kühlemeier, H., Van Til, A., Feenstra, H., & Hemker, B.(2013). Balans van de schrijfvaardigheid in het basis- en speciaal basisonderwijs 2. Periodieke Peiling van het Onderwijsniveau (No. 53) [Present state of writing competency in elementary and special education 2. Periodical assessment of the level of education]. Arnhem: Cito.

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

*Mason, L. H., Davison, M. D., Hammer, C. S., Miller, C. A., & Glutting, J. J. (2012). Knowledge, writing, and language outcomes for a reading comprehension and writing intervention. *Reading and Writing*, *26*(7), 1-26.

Morphy, P., & Graham, S. (2012). Word processing programs and weaker writers/readers: A meta-analysis of research findings. *Reading and Writing, 25,* 641-678. http://dx.doi.org/10.1007/s11145-010-9292-5

National Commission on Writing (2003). *The need for a writing revolution. The neglected "R"*. New York: College Entrance Examination Board.

Northwest Regional Educational Library (2013). *About 6 + 1 Trait Writing.* Retrieved September 2, 2013, from http://www.educationnorthwest.og/resource/949.

*Olson, V. L. B. (1990). The revising processes of sixth-grade writers with and without peer feedback. *The Journal of Educational Research*, *84,* 22-29.

*Paquette, K. R. (2008). Integrating the 6+ 1 Writing Traits Model with Cross-Age Tutoring: An Investigation of Elementary Students' Writing Development. *Literacy Research and Instruction*, *48*(1), 28-38. http://dx.doi.org/10.1080/19388070802226261

*Puma, M., Tarkow, A., & Puma, A. (2007). The Challenge of Improving Children's Writing Ability: A Randomized Evaluation of" Writing Wings". Institute of Education Sciences. Retrieved September 2, 2013 from: http://eric.ed.gov/?id=ED504279.

*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.

Rijlaarsdam. G., Bergh, H. van den, Couzijn, M. Janssen, T., Braaksma, M., Tillema, M. Steendam, E. van, Raedts, M. (2011). Writing. In, Graham, S., Bus, A. Major, S. & Swanson, L. (Eds.), *Application of Educational Psychology to Learning and Teaching.* APA Handbook Volume 3, 189-228.

Rogers, L., & Graham, S. (2008). A meta-analysis of single subject design writing intervention research. *Journal of Educational Psychology, 100*(4), 879-906. http://dx.doi.org/10.1037/0022-0663.100.4.879

*Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (1999). Effects of self-evaluation training on narrative writing. *Assessing Writing*, *6*(1), 107-132. http://dx.doi.org/10.1016/S1075-2935(99)00003-3

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

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

Schoonen, R. (2005). Generalizability of writing scores: An application of structural equation modelling. *Language Testing*, 22, 1-30. http://dx.doi.org/10.1191/0265532205lt295oa

*Schunk, D. H., & Swartz, C. W. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology*, *18*(3), 337-354. http://dx.doi.org/10.1006/ceps.1993.1024

Smits, M. (2009). *Schrijven en leren op de pabo*. *[Writing and learning in teacher training]* Doctoral dissertation. Nijmegen: Radboud University.

Stoeldraijer, J. (2012). *Kwaliteitskaart onderwijs in het schrijven van teksten.[Quality card education in writing texts]* The Hague: School aan Zet.

*Tienken, C. H., & Achilles, C. M. (2003). Changing Teacher Behavior and Improving Student Writing Achievement. *Planning and Changing*, *34*(3), 153-168.

*Torrance, M., Fidalgo, R., & García, J. N. (2007). The teachability and effectiveness of cognitive self-regulation in sixth-grade writers. *Learning and Instruction*, *17*(3), 265-285. http://dx.doi.org/10.1016/j.learninstruc.2007.02.003

*Varble, M. E. (1990). Analysis of writing samples of students taught by teachers using whole language and traditional approaches. *The Journal of Educational Research*, 245-251. http://dx.doi.org/10.1080/00220671.1990.10885965

*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*(4), 757-784. http://dx.doi.org/10.1016/j.cedpsych.2007.12.004

*Yarrow, F., & Topping, K. J. (2001). Collaborative writing: The effects of metacognitive prompting and structured peer interaction. *British Journal of Educational Psychology*, *71*(2), 261-282. http://dx.doi.org/10.1348/000709901158514

Appendix: Forest plot of expected (grey) and observed effect size, and 95% confidence interval per study