



Write, talk and rewrite: the effectiveness of a dialogic writing intervention in upper elementary education

Renske Bouwer¹ · Chiel van der Veen^{2,3}

Accepted: 28 July 2023 / Published online: 30 August 2023
© The Author(s) 2023

Abstract

In this research, we developed and empirically tested a dialogic writing intervention, an integrated language approach in which grade 5/6 students learn how to write, talk about their writing with peers, and rewrite. The effectiveness of this intervention was experimentally tested in ten classes from eight schools, using a pretest–posttest control group design. Classes were randomly assigned to the intervention group (5 classes; 95 students) or control group (5 classes; 115 students). Both groups followed the same eight lessons in which students wrote four argumentative texts about sustainability. For each text, students wrote a draft version, which they discussed in groups of three students. Based on these peer conversations, students revised their text. The intervention group received additional support to foster dialogic peer conversations, including a conversation chart for students and a practice-based professional development program for teachers. Improvements in writing were measured by an argumentative writing task (same genre, but different topic; near transfer) and an instructional writing task (different genre and topic; far transfer). Text quality was holistically assessed using a benchmark rating procedure. Results show that our dialogic writing intervention with support for dialogic talk significantly improved students' argumentative writing skills ($ES = 1.09$), but that the effects were not automatically transferable to another genre. Based on these results we conclude that a dialogic writing intervention is a promising approach to teach students how to talk about their texts and to write texts that are more persuasive to readers.

Keywords Writing education · Revision · Dialogic talk · Peer feedback · Primary education

✉ Renske Bouwer
r.bouwer@uu.nl

¹ Institute for Language Sciences, Utrecht University, Trans 10, 3512 JK Utrecht, Netherlands

² LEARN! Research Institute, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

³ Windesheim University of Applied Sciences Almere, Zwolle, Netherlands

In primary schools, only little teaching time is devoted to the productive language skills of writing and speaking (Inspectorate of Education, 2019, 2021). Also, writing, reading and speaking are often covered in separate lessons, making it difficult for students to use language in a variety of meaningful contexts as a means of communication (Sperling, 1996). This is particularly problematic for the teaching of writing, as writing is a social act—writers not only write for a general purpose, but they also particularly write to communicate ideas to others (Graham, 2018). Demonstrating audience awareness before, during and after writing is therefore considered to be one of the most important aspects of writing proficiency (Rijlaarsdam et al., 2009). However, beginning writers do not automatically take the perspective of the reader into account while writing and afterwards, they hardly revise their text accordingly (Bereiter & Scardamalia, 1987; Kellogg, 2008). As a result, the majority of students in primary school are unable to write texts in which they successfully convey their message to a reader (Inspectorate of Education, 2021; National Center for Education Statistics, 2012).

Based on previous research (see for example Casado-Ledesma, 2021), we argue that fostering a direct dialogue between writers and readers can support students to become more aware of their audience, and hence, support them in writing texts in which they communicate their ideas more effectively. For this purpose, together with educational practitioners, we developed a dialogic writing intervention, which is an integrated approach to written and oral language skills in primary education. This intervention consists of three steps: (1) writing a first draft of an argumentative text, (2) engaging students in small group dialogue about their own written texts, and (3) revising their text based on these conversations. The effectiveness of this writing intervention for fifth- and sixth-grade students (aged 9–13 years) was empirically tested in this study.

Developing audience awareness through dialogic peer talk

From a sociocultural perspective, writing and speaking have a similar function: to convey information to a recipient through language (Sperling, 1996). However, where in spoken language there is a direct dialogue between speaker and hearer, the dialogue between writer and reader is indirect. This implies that writers have to anticipate whether their text has the intended communicative effect on the reader and adjust their text accordingly. Experienced writers do this by consistently taking a reader's perspective during the writing process (Kellogg, 2008). Putting themselves in the reader's shoes is, however, very difficult for beginning writers; they write as they talk, resulting in "and then, and then"-like stories (Bereiter & Scardamalia, 1987). They also hardly revise their text; once the first ideas are on paper, the text barely changes and the only revisions that they make are superficial (Chanquoy, 2001). This is particularly problematic when writing argumentative texts, as in this genre revision is an important process to ensure the text becomes as convincing as possible. For example, for argumentative text it is essential to critically re-read the text from the perspective of the reader, in order

to determine whether the arguments are convincing enough and what substantial changes in the text are needed to persuade the reader more effectively.

Fostering a dialogue between writer and reader, combined with strategies before and after writing, can support elementary students to write persuasive texts (Casado-Ledesma, 2021; Rijlaarsdam et al., 2009). In addition, meta-analyses show strong effects of peer interaction on text quality (Graham et al., 2012; Koster et al., 2015). In these peer conversations, writers themselves experience how their text comes across to readers, which provides them with concrete suggestions for improving the text. These evidence-based instructional practices have been translated into the strategy-based writing method *Tekster* for upper-elementary students (Bouwer & Koster, 2016). Two large-scale intervention studies showed that after four months of strategy-focused instructions, the writing skills of students aged 10–12 improved by one-and-a-half grade (Bouwer et al., 2018; Koster et al., 2017). Despite these positive results, classroom observations showed that students still found it difficult to have meaningful conversations about each other's texts, taking the reader's perspective and providing the other with substantive feedback.

That peer conversations are hardly ever about arguments in the text was also seen in a small-scale pilot study with elementary students prior to the current research project (Bodewitz, 2020). Again, without any instructional support, peer conversations about students' own written texts turned out to be mostly superficial and mainly focused on the formal aspects of the text, such as spelling, punctuation and layout. Also, the conversations were hardly dialogic in nature; students did not ask each other questions about the ideas in the text and rarely deepened each other's contributions. Not surprisingly, they made almost no meaningful changes when revising their text; their revisions focused mainly on superficial aspects of the text, such as language errors or layout. This raises the question of how we can support students in deepening their conversations about each other's texts.

Research on classroom conversations shows that dialogic conversations in which teachers ask open-ended questions and students are challenged to take the other's perspective, reason, think together and listen critically (Hennessy et al., 2016) have a positive effect on students' reasoning skills, motivation and domain-specific knowledge (Dobber, 2018; Resnick et al., 2015; Van der Veen, 2017). Such dialogic conversations also have a positive effect on students' communication skills (Van der Veen et al., 2017). However, there is hardly any scientific research on the use of dialogic conversations in the context of writing instruction (Fisher et al., 2010; Herder et al., 2018). Also, existing research tends to be correlational and focused on how teachers employ dialogic conversation techniques during whole-class conversations (cf. Myhill & Newman, 2019). For example, a recent study shows that teachers' asking open-ended questions and questioning is correlated with higher text quality for argumentative writing (Al-Adeimi & O'Connor, 2021). Yet, little is known about what support students need to have dialogic conversations with each other about their own written texts and how this kind of dialogic peer talk improves students' writing skills.

Research aims

In this research project, we build on the knowledge that already exists about learning to engage in dialogic conversations and explore the extent to which a dialogic writing intervention can contribute to the improvement of writing skills. More specifically, we were interested in the following research questions: what is the effect of a dialogic writing intervention on students' writing improvements in (a) argumentative writing between pre- and posttest (same genre, near transfer; RQ1), and (b) instructional writing on the posttest (other genre, far transfer; RQ2). In addition, we examined whether (c) possible effects of the intervention on students' writing performance at posttest can be explained by level of writing proficiency at pretest (RQ3). Finally, we investigated the fidelity and social validity of our dialogic writing intervention (RQ4).

Based on cognitive writing process theories by Flower and Hayes (1981) and Kellogg (2008) as well as sociocultural models that emphasize the importance of the reader for writing (Graham, 2018), we hypothesized that our dialogic writing intervention will support students with different levels of writing proficiency in becoming more aware of the intended reader and, as a result, write texts that communicate messages more effectively. With this research, we aim not only to optimize interventions for teaching writing, but also to develop more generic knowledge about how to successfully integrate different language domains such as speaking, reading and writing into classroom practice and what this requires of teachers and students.

Method

Ethics

The study obtained ethical approval from the Faculty Ethics Review Committee Humanities (FETC-GW, reference number 21–147-03). Teachers and school leaders of the participating schools were informed about the purpose and procedure of the study prior to the study and were asked to consent to data collection in the classrooms. Parents and students received the same information about the study and were asked to return signed consent forms back to the teacher. The students who did not receive permission to participate did attend the classroom program and completed the corresponding assignments as part of regular language classes, but these data were not stored and included in the data analysis. The background information of the students with permission and the written texts were stored encrypted on a secure server at the university in such a way that the data could not be traced back to individual students. The recordings of conversations and lessons have been transcribed, with any names or details traceable to individuals removed from the transcripts. The data are kept securely for 10 years and can only be accessed by involved researchers.

Participants

Teachers from ten grade 5/6 classes of eight schools and their 278 students participated in this intervention study. We received consent to participate in the study from 212 students and their parents, which is 76% of all students. Teachers and their students were randomly assigned to either the intervention group (5 teachers; 95 students) or the control group (5 teachers; 115 students). With this sample, the power is at least 0.80 to ascertain small effects in a multilevel model in which pupils are nested in classes. Specifically, an a priori power analysis for multilevel modeling (see Hox, Moerbeek, & van de Schoot, 2018) revealed that an effective sample size of 151 students was needed to ascertain a small effect with a power of 0.81, based on two measurement occasions, an average class size of 20 and an intraclass correlation of 0.10.

The total distribution of girls and boys in the sample was nearly equal, with 47% girls and 53% boys. The age of the students ranged from 9 to 13 years, with a mean of 10.68 years ($SD=0.73$). The majority of the students (67%) were from grade 6. From only three-quarters of the students, we were able to collect additional information about their home language. This showed that a quarter of the pupils (35 out of 141 pupils) spoke another language at home besides Dutch, including Turkish ($n=7$), Arabic ($n=2$), Berber ($n=2$), English ($n=4$), Hungarian ($n=2$), Cape Verdean ($n=2$) and Polish ($n=1$). There were no differences in student background between conditions.

Procedure and materials

The effect of the intervention was experimentally tested at the beginning of school year 2021–2022 using a pretest–posttest control group design. Both groups followed the same eight-week lesson program for dialogic writing, but the intervention group received additional support for dialogic conversations, including a conversation chart for students and a Practice-Based Professional Development program for teachers (PBPD; Harris et al., 2023). By implementing the same writing lesson program in the control group, instead of using a business-as-usual control group, we ensured that students in both conditions spent the same amount of time and attention on writing, discussing texts and rewriting. Also, teachers were not aware of the existence of two different conditions.

Lesson program for dialogic writing in the control and intervention group

Together with experienced primary school teachers and teacher trainers, we have developed a dialogic writing program consisting of eight lessons – one lesson each week. This lesson program was assigned to all participating classes, in both the intervention and control condition. During this lesson program, students wrote four argumentative texts about sustainability: (1) a personal experience, (2) an argumentative text, (3) a persuasive letter, and (4) a reply to a letter. Sustainability is a meaningful

theme that is urgent for students; after all, they are the next generation living on this planet. The four texts students had to write during the lesson series built on each other in terms of content and audience, in which taking the reader's perspective becomes increasingly complex. That is, in lessons 1 and 2 they began writing a text based on a personal experience in which they give an example of something they already do to live sustainably, such as reusing clothes or collecting plastic waste. The second text they wrote in lessons 3 and 4 is an argumentative text in which they presented factual arguments to defend a certain position related to sustainability. Besides, they backed it up with arguments from source texts. They also learned to rebut possible counter arguments from peers. The third text students wrote in lessons 5 and 6 was a persuasive letter to a yet unknown student from another school. The goal was to persuade the person with personal and factual arguments to implement their sustainable idea. To create a meaningful context, the students actually sent the letters to students from another school. The final text students wrote in lessons 7 and 8 was a response to the letter they received from someone else. Students read the letter and discussed whether they were convinced by the other person's sustainable idea and why. Based on that discussion, they wrote a letter back.

The four texts are characterized by the communicative aspect of writing: they are written to convince a reader to also take action for a sustainable planet. To help students achieve this goal, we used a similar format in all of the lessons. In the first lesson for each new text, students gathered ideas for their text by reading source texts and discussing their ideas with peers. They incorporated those ideas into a first version of their text. In the following lesson, they shared and discussed their texts in groups of three, after which students revised their text individually. By talking with readers before, during and after writing, they learned which arguments already work well and where they needed to add more information or additional arguments in the text to convince the reader.

Intervention group: additional support for dialogic talk about writing

To further deepen students' conversations about each other's texts, and to make them more dialogic, we added additional support for both students and teachers in the dialogic writing intervention condition. To do so, we developed a conversation chart with examples of various dialogic conversation techniques, such as asking open questions, deepening each other's contributions, critical listening and summarizing, see also Appendix 1. The conversation chart consists of the following four steps:

1. The writer reads her/his own text while the other two students listen and think about what the writer wants to say with the text.
2. The writer starts the conversation with the questions from the conversation card, such as whether the text is clear and convincing.
3. Readers respond to the writer's questions and are encouraged to respond to each other and to ask follow-up questions, so that students explain their feedback and deepen the conversations.
4. The writer summarizes what feedback she/he would like to work with in revising the text, to establish the link between the conversation and the rewriting process.

Next to support for the students, the teachers in the intervention group received an intensive PBPD over the course of the intervention to adequately implement the principles of dialogic writing in their classrooms. This program consisted of a three-hour workshop and a manual prior to the intervention that provided teachers that participated in the intervention condition with an introduction on the theory behind dialogic writing and how to use and promote dialogic conversations with students. This included techniques for how to invite students to share their ideas by asking open-ended questions and providing time for students to think, as well as to promote peer discussion by connecting the ideas of students and asking follow-up questions to deepen the conversation. In addition to these conversation techniques, teachers learned how to focus on the communicative effectiveness of the text by discussing the most important criteria for argumentative texts, such as well-supported arguments, a clear line of reasoning and good choice of words. The teachers observed and discussed videos in which another teacher modeled how to apply these principles of dialogic writing during a writing lesson. There were also exercises in which teachers could experience for themselves the benefits of dialogic conversation for rewriting their own text. Finally, they received six basic conversation rules for promoting good peer conversations in an open, safe and respectful setting, such as listening to each other and really trying to understand each other by asking questions and responding to each other, see also Appendix 2.

In addition to the workshop, teachers received sustained support during the intervention period. This included individual coaching for lessons 2 and 6 in their own classroom context from one of the teacher trainers from the consortium. These trainers observed lessons 2 and 6 in the classroom and afterwards reflected with the teachers on how they implemented the principles of dialogic writing in practice. After lessons 4 and 7 there was an online training meeting with all participating teachers and teacher trainers, during which teachers discussed with each other what went well and what they found difficult. They also reflected on each other's lessons using short video clips from their own practices.

Measuring fidelity of implementation and social validity of the lesson program

To investigate whether teachers implemented the dialogic writing lessons into their classroom practice as intended, we focused on five key elements of fidelity as suggested by Smith et al. (2007): design, PBPD, intervention delivery, intervention receipt, and intervention enactment. First, the *design* of our intervention was developed in close collaboration with teachers, and piloted in an earlier study (Bodewitz, 2020). Furthermore, we made sure that the goals of our writing lessons covered the goals of Dutch primary schools. We asked all participating teachers in the control and intervention condition to keep a logbook for each lesson to investigate the extent they were able to conduct all lessons according to our teacher manual. They were asked to provide the following information about each lesson: preparation time, lesson duration, appreciation of the lesson (on a scale from 1 to 10; 1=very low, 10=very high), aspects that went well, aspects that went less well, and whether they made any modifications or adjustments to the lesson. Teachers also shared all draft

and revised versions of the texts that were written during the lessons, as well as audio recordings of the peer conversations and classroom instructions.

Our *PBPD* was delivered by experienced teacher trainers that were part of our consortium. All teachers in the intervention condition received the same *PBPD* to maximize fidelity of *intervention delivery*. Furthermore, during the intervention the teacher trainers and research team had regular meetings to discuss intervention delivery. Afterwards, we conducted interviews with all participating teachers to get in-depth insight in their experiences with implementing the principles of dialogic writing in the classroom, as well as their satisfaction with the developed teaching materials and the additional conversational support (only for teachers in the intervention condition). This provides us with information on the *intervention receipt* and social validity of the lesson program for dialogic writing. Interviews were conducted by the first author, and they were video-taped and transcribed.

Finally, to evaluate *intervention enactment*, logbooks, transcriptions of post-interviews, text written by students, and audio recordings of peer conversations and classroom instructions were examined and checked against our intervention materials (e.g., teacher manual, workshop, etc.). Although teachers were forced to make some changes due to school closings and time constraints, in general no abnormalities were found. However, not all teachers were able to complete all steps of the lessons, see Results for more details on intervention enactment.

Pre- and post-measurement of text quality

To measure the effects of the dialogic writing intervention on writing skills, students wrote a persuasive text before and after the lesson series on a topic that was different from the focus of the lesson series. As in the lesson series, the goal was to persuade peers, this time not about a sustainable idea, but about a pet for the classroom (pre-test) and about an outing to an amusement park with the whole class (post-test). Afterwards, they also wrote an information letter to investigate whether the effects of the intervention were transferable to another genre. In this text, they gave advice on how to write a good letter to a fictional student coming to the Netherlands next year. These writing tasks were already developed, tested, and used to measure writing progress in two large-scale intervention studies for students in comparable grades (Bouwer & Koster, 2016).

The quality of students' written texts at the pre- and posttest was assessed holistically and comparatively based on its communicative effectiveness by juries of three independent raters using a benchmark rating scale with five benchmark texts of increasing text quality (cf. Bouwer & Koster, 2016; Bouwer, Van Steendam, & Lest-erhuis, 2023). The scale on which the benchmarks are placed can be considered as an interval scale. The benchmark in the center of the scale is a text of average quality for students in upper-elementary grades, which is assigned an arbitrary score of 100 points. The other texts on the scale are one and two standard deviations below average, with a score of respectively 85 and 70 points, and one and two standard deviations above average, with a score of 115 and 130 points. Raters had to compare each student text to the five benchmarks on the scale and score it accordingly. These scores could theoretically range from 0 to infinite, however, in practice the scores

across the three writing tasks ranged from 57 to 141 points. This corresponds to three standard deviations below and above average. We used the same benchmark scale for the persuasive writing task at the pre- and posttest, as previous research demonstrated that the same benchmark rating scale can be used reliably and validly for rating texts in the same genre but on a different topic (Bouwer, Koster, & Van den Bergh, 2023). For the informative writing task, we used a different benchmark scale. Both benchmark scales were validated in previous research (Bouwer & Koster, 2016).

There were nine raters in total, who were all experienced teachers. They received a short training in which they learned to use the benchmark scales. After this, they independently rated three subsets of all the texts, blind to condition, using a prefixed design of overlapping rater teams, each text was rated by three raters. This overlapping rater design also allows to approximate the reliability of both individual and jury raters (Van den Bergh & Eiting, 1989). The average reliability of jury ratings across the three tasks was high, with $\rho=0.88$ for the first writing task, $\rho=0.86$ for the second writing task, and $\rho=0.93$ for the third writing task. The final text quality score for each text was determined by the average score of the three raters.

Data-analysis

The data of the present study are hierarchically structured. That is, writing scores (level 1, $N=530$) are nested within students (level 2, $N_s=212$), and students are nested within classes (level 3, $N_c=10$). To take this nested structure into account, multilevel modeling was applied, using linear mixed model analyses with maximum likelihood (ML) estimations in SPSS (version 27) following the procedures of Snijders and Bosker (2012). In the multilevel models all students, including those with partly missing values, are taken into account. In total, percentages of missing data ranged from 11.8% for the pretest to 13.2% for the first posttest and 25.0% for the second posttest. Only 8.5% of the students missed scores for both posttest measures. Little's MCAR test also revealed that the missing values were missing at random for the three measurement occasions ($\chi^2(9)=14.15, p=0.12$).

In the random intercept-only model (Model 0) we estimated the ICC as an indication of the proportion of variance that can be attributed to classes and students. To test the effects of the intervention on students' writing performance, five multilevel models were compared in which parameters were added systematically. Model 1 is the homoscedastic model in which we held the variance between measurement occasions constant. This model served as a baseline to which we compared the more comprehensive models. In this basic model we accounted for random effects between classes (S^2_c) and between and within students (respectively S^2_s and S^2_e). That is, scores on the main variables were allowed to vary between and within students, and between classes. In Model 2, measurement occasions were considered as a repeated measure (unstructured), in which text quality scores were allowed to vary between measurement occasion. In Model 3, condition (i.e., intervention group versus control group) was added to investigate the main effect of additional support for students and teachers (e.g., PBPD and conversation chart). Next, in Model 4, we

added the interaction effect between condition and measurement occasion, to investigate the effect of the intervention over time. Finally, two control variables were added as fixed effects: gender and grade. Models were compared using the log-likelihood ratio tests for model improvement (alpha of 0.05).

Results

Effect of the dialogic writing intervention on text quality

The random intercept-only model (Model 0) confirmed that the text quality scores are clustered within students and classes. In particular, the total variance in text quality scores was 226.73, with considerable variance between classes and students of respectively 28.24 and 59.06. Hence, 13% of the total variance in text quality is attributable to classes ($ICC=0.13$).

Table 1 shows the results of the fit and comparisons of the planned models 1–5 as well as the parameter estimates for each model. As can be seen, the heteroscedastic model in which the variances were allowed to differ between measurement occasion fitted the data better than the homoscedastic model (Model 2 versus Model 1, $\chi^2(4)=7.20$, $p<0.01$), indicating that variance in text quality scores between students differed between measurement occasions. In fact, parameter estimates show that the variance between students decreased between the first and second measurement occasion, indicating that students' writing became more homogeneous after the lesson program for a different writing task in the same genre (near transfer). The variance between students increased again for the third measurement occasion, in which students wrote a text in a different genre (far transfer). Further, the parameter estimates show that text quality scores covaried between measurement occasions, indicating a positive correlation between the three writing tasks.

Results of the model comparison further show that there was no main effect of condition (Model 3 versus Model 2, $\chi^2(1)=2.26$, $p=0.13$), indicating that the average writing scores were the same for students in the two conditions. There was, however, an interaction effect between condition and measurement occasion (Model 4 versus Model 3, $\chi^2(3)=18.96$, $p<0.001$). This means that, while taking into account the variance between classes and students, text quality scores measured at two occasions were not the same for students in the intervention condition and the control condition. The differences in text quality scores between condition and measurement occasion are presented in Table 2 below. The interaction effect was also apparent when controlling for the effects of gender and age (Model 5 versus Model 4, $\chi^2(2)=48.59$, $p<0.001$). In particular, average writing scores were 4.18 points ($SE=1.98$) lower for students in grade 5 than for students in grade 6 ($t(159.94)=-2.11$, $p=0.04$), and girls scored on average 8.89 points ($SE=1.34$) higher than boys ($t(191.56)=6.65$, $p<0.001$). This final model explained 24% of the total variance in text quality scores between classes ($R_2^2=0.24$). The explained variance in text quality scores between students was 9% on the pretest, 26% on the first posttest and 11% for the second posttest.

Table 1 Parameter estimates and goodness of fit for planned multilevel models of text quality (N = 212)

	Model 1	Model 2	Model 3	Model 4	Model 5
Fixed effects					
Occ 1 (pretest)	93.51 ^{***} (1.86)	93.55 ^{***} (1.81)	91.12 ^{***} (2.20)	88.50 ^{***} (2.29)	84.25 ^{***} (2.51)
Occ 2 (post, near transfer)	106.67 ^{***} (1.86)	106.72 ^{***} (1.74)	104.31 ^{***} (2.16)	106.29 ^{***} (2.19)	102.00 ^{***} (2.44)
Occ 3 (post, far transfer)	101.25 ^{***} (1.89)	101.07 ^{***} (1.82)	98.60 ^{***} (2.22)	98.24 ^{***} (2.32)	93.98 ^{***} (2.52)
Control group			4.81 (2.95)		
Control group * occ 1				9.64 [*] (3.19)	12.05 ^{**} (3.38)
Control group * occ 2				1.02 (3.07)	3.35 (3.29)
Control group * occ 3				5.56 (3.28)	8.08 [*] (3.43)
Grade = 5					-4.18 [*] (1.98)
Gender = girl					8.89 ^{***} (1.34)
Random effects					
Classes (S^2)	25.65 (14.00)	22.73 (12.95)	16.42 (10.34)	16.17 (10.22)	19.50 (11.35)
Students (S^2)	73.66 (11.59)				
Occ 1	186.29 (20.24)	0.33	0.46	0.33	0.45
Occ 2	53.45 (13.00)	138.92 (15.40)	52.96 (12.92)	141.99 (16.14)	0.57
Occ 3	83.59 (15.84)	87.61 (14.80)	179.57 (21.06)	90.04 (15.44)	181.22 (21.53)
				181.22 (21.53)	67.53 (12.83)
				90.75 (15.27)	123.61 (13.66)
				80.74 (15.36)	52.66 (13.02)
					149.10 (18.04)

Table 1 (continued)

	Model 1	Model 2	Model 3	Model 4	Model 5
Error variance (S^2_e)	93.87 (7.33)				
Goodness of fit					
AIC	4164.85	4157.65	4157.39	4144.69	4100.10
BIC	4190.49	4200.38	4204.39	4200.24	4164.16
-2Loglikelihood	4152.85	4137.65	4135.39	4118.69	4070.10
$\Delta\chi^2$		7.20**	2.26	18.96***	48.59***
Δdf	6	4	1	3	2

Standard errors are in parentheses. All p -values in this table are two-tailed

* $p \leq .05$. ** $p \leq .01$. *** $p \leq .001$

Table 2 Estimated means for text quality scores per condition on pre- and posttest measures

	Control condition				Intervention condition			
	<i>M</i>	<i>SE</i>	95% CI		<i>M</i>	<i>SE</i>	95% CI	
Pretest	98.66	2.34	93.53	103.78	86.60	2.47	81.29	91.92
Posttest (near transfer)	107.70	2.28	102.66	112.74	104.35	2.40	99.14	109.55
Posttest (far transfer)	104.34	2.40	99.17	109.51	96.33	2.48	91.04	101.63

Table 2 shows the estimated marginal means for the text quality scores for each condition and measurement occasion, adjusted for the other variables in the model. To verify the direction of the interaction effect between condition and measurement occasion, three specific contrasts were performed on the data. Results of the first contrast on the interaction effect showed that the writing performance of students in the intervention group improved more strongly between pretest and the first posttest compared to students in the control group ($t(188.14)=4.10$, $p<0.001$), with a significant increase in text quality of 17.79 points from pre- to posttest for students in the intervention group ($SE=1.54$, $p<0.001$) and 9.17 points for students in the control group ($SE=1.43$, $p<0.001$). The magnitude of this effect was estimated by comparing the effect of the intervention to the total amount of variance pooled over measurement occasion (Cohen's *d*). This resulted in an estimated effect size in the intervention condition of 1.09, while generalizing over students and classes. These results show that students improved their writing performance in argumentative text between pre- and posttest (near-transfer; RQ1).

Results of the second contrast showed that improvements did not transfer to a different genre: text quality scores on the second posttest that measured far transfer effects were lower compared to text quality scores on the first posttest (same genre, near transfer; $t(166.88)=-2.43$, $p<0.05$). This was the case for the students in the intervention group who scored 9.74 points lower at the second posttest ($SE=1.56$, $p<0.001$), as well as for the students in the control group who scored 3.51 points lower at the second posttest ($SE=1.54$, $p=0.009$). These results show that students did not transfer their improved writing performance to another genre (far transfer; RQ2).

There was, however, no complete decline in writing performance at the second posttest as was shown by the results of the third contrast. That is, the differences between the two conditions at the beginning of the intervention were no longer apparent at the second measurement occasion ($t(177.17)=4.08$, $p=0.06$). More specifically, whereas students in the control group scored 9.64 points higher at the pretest than students in the intervention condition ($SE=3.19$, $p=0.01$), this difference decreased to 5.56 points at the second posttest which was a non-significant difference ($SE=3.28$, $p=0.11$). Thus, particularly for students in the intervention group, some learning gains seem to have remained. Figure 1 illustrates the differences between conditions for the three different measurement occasions (pretest, near transfer, far transfer).

Aptitude treatment interaction

As there were large differences between students at the pretest, we investigated in a separate analysis whether the effect of the intervention depended on students' writing proficiency (RQ3). For this analysis, we estimated the regression of the pretest scores on the scores of the posttests, and its interaction with condition. For the first posttest (near transfer), it was shown that pretest scores significantly predicted scores at the posttest ($\beta=0.34$, $p=0.001$), but there was no significant interaction effect of condition by text quality scores at pretest ($\beta=0.62$, $p=0.16$). The results for the second posttest (far transfer) were comparable, with a significant regression of scores at pretest on posttest ($\beta=0.54$, $p<0.001$), but again no significant interaction between pretest scores and condition ($\beta=-0.18$, $p=0.69$). These results show that students' writing performance at the posttest is explained by their level of writing proficiency at the posttest, but that there is no interaction between condition and pretest scores.

Implementation fidelity and social validity of dialogic writing lessons in the classroom

We used the teacher logbooks and interviews to gain further insight into whether the intervention was implemented as intended (e.g., intervention enactment), how satisfied the teachers were with the intervention and what support teachers found particularly helpful as well as the things they struggled with (RQ4). Together, this provides us with information regarding intervention fidelity, and the internal and social validity of the dialogic writing intervention.

First, the logbooks and student products showed that most teachers completed all lessons, but in both the intervention and control group there were two classes who were not able to complete all the steps of the final two lessons in which they had to write a response to a letter. That the teachers were not able to complete this writing task was because they had to switch to homeschooling due to covid. Although they tried to finish the lessons at home, this turned out to be too challenging, especially regarding the final steps that included the peer dialogue and text revision.

Second, teachers indicated in the logbooks that the topic of sustainability was interesting for many students, but that they also found it a difficult topic to write about. It is therefore important to select a topic or theme that students want and can write about. Background information on the topic can help teachers select appropriate resources and examples for the first lesson. Furthermore, the interviews revealed that the design of the lesson series was slightly more difficult for teachers who are not used to working with themes in their schools. They indicated that they found it difficult to keep their students motivated to work on the same theme for eight weeks. According to them, the principles of dialogic writing (writing—talking—rewriting) are also suitable for shorter writing assignments on different topics.

Third, the logbooks revealed that in most classes the lessons took longer than estimated. On average, the lessons lasted 78 min, ranging from 45 to 120 min. The

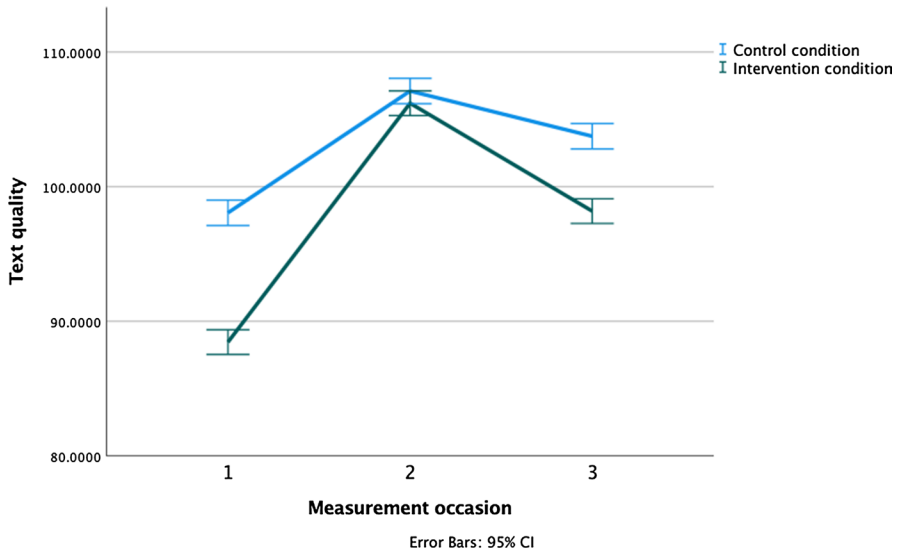


Fig. 1 Differences in text quality scores between intervention group (green line) and control group (blue line) for the different measurement moments. Measurement occasion 1 is the pretest, measurement occasion 2 is the first posttest (same genre; near transfer) and measurement occasion 3 is the second posttest (other genre; far transfer). (color figure online)

preparation time for teachers was 34 min per lesson, ranging from 15 to 75 min. The lesson duration was particularly longer for lesson 2, 4, 6, and 8 in which students engaged in conversation with each other and then had to rewrite their text, but teaching time was also longer for lesson 3 in which students had to use arguments from source texts to support their own point of view. In the interviews, teachers indicated that more time should be set aside for reading sources and selecting appropriate arguments to support one's own opinion. There was also one teacher who indicated that the planned 20 min writing time is too short for the students to finish one draft, even after 40 min of instruction and brainstorming for ideas.

Fourth, the interviews with teachers revealed that the most challenging step for students was to engage in a substantive dialogue about their own texts. Students appeared to find this very difficult, especially at the beginning of the lesson series. Teachers also indicated that they find it difficult to provide sufficient support in this process and that they would prefer to be present at every conversation to be able to adjust, but that this is not practically feasible. The teachers in the intervention group indicated that they benefited a lot from the training. They also argued that the conversational chart for students is essential, as it stimulates the students to deepen their conversations.

Finally, despite the challenge of encouraging good conversations in the classroom, all teachers reported that as the lesson series progressed and students practiced with the cycle of writing—talking—rewriting a few times, that they were becoming more critical of each other's texts in their conversations. Whereas at the beginning students had to get used to discussing the ideas in their texts with each other, by the

end of the lessons the process of dialogic writing was well ingrained. The teachers indicated afterwards that most students were actively and seriously discussing and improving their texts. They were also especially enthusiastic about the final lessons in which they exchanged their letters with those from another school. Students really enjoyed reading others' letters and writing something back. Most teachers also indicated that they would like to continue implementing the principle of dialogic writing in writing instruction. Some have even already successfully applied it to other writing tasks outside of the planned intervention, for instance by writing invitations to parents for the final musical or writing poems (see Kooijman, 2022).

Discussion

The purpose of this study was to investigate the extent to which a dialogic writing intervention can improve students' writing skills. Our research questions were: what are the effects of a dialogic writing intervention on students' writing improvements in (a) argumentative writing between pre- and posttest (same genre; near transfer; RQ1), (b) instructional writing on the posttest (other genre; far transfer; RQ2). In addition, we studied whether (c) possible effects of the intervention on students' writing performance at posttest can be explained by level of writing proficiency at pretest (RQ3). Finally, we investigated the fidelity and social validity of our intervention (RQ4). The results show a positive picture. That is, the dialogic writing intervention with additional support for students and teachers supports students to have dialogic conversations about each other's own written texts and revise their text accordingly. In these conversations, students ask each other more questions and give each other more feedback on the content of the texts. More importantly, the dialogic writing intervention has a positive effect on students' writing performance ($ES = 1.09$). After attending eight lessons, students write texts that are more persuasive and of higher quality. This is particularly true for the genre in which they practiced (i.e., argumentative writing; RQ1), and to a lesser extent for a new text genre (i.e., informative writing; RQ2). Furthermore, our results show that students' posttest writing performance can be explained by their writing proficiency at the pretest, although our intervention did not affect this relation (RQ3). Finally, fidelity and social validity of our dialogic writing intervention were high (RQ4).

Based on this study, we can conclude that by taking a systematic approach to write, talk, and rewrite, students develop audience awareness, which over time leads to improved writing quality. Already with a little help and practice, students can have dialogic conversations about each other's texts. With this, we were able to successfully integrate the various language domains into a lesson series on the topic of sustainability. After all, students were not only writing texts, but they also read each other's texts and source texts for additional background information about their own text, and they talked about it with each other. Follow-up research should reveal

whether and in what way a dialogic writing intervention also contributes to students' reading and speaking skills.

Limitations

A limitation of the current study is that only one text per genre was measured, making it difficult to generalize the effects to writing skills in general (Bouwer et al., 2015). In addition, the writing assignments were taken immediately after the lesson series, therefore it remains unclear whether there is also a long-term effect of the dialogic writing intervention. As a result, we can only show how students performed on these tasks immediately after the lessons. However, the topic of the pre- and post-measurement did differ from the topic that was the focus during the lessons. This does raise the expectation that after the lesson series, students will also score higher for text quality for other texts of a similar genre. Follow-up (longitudinal) research with more tasks and multiple measurement times is necessary to confirm this expectation.

Another limitation of the current study was the timing of the implementation of the lesson series. The second half of the lessons fell in the middle of a covid fall wave, requiring some classes to switch to homeschooling. Because it was necessary for the study to keep the moments of pre- and post-measurement the same in all classes, we chose to continue the lessons online. However, it proved to be difficult for teachers to properly supervise students' discussions and revisions online. Also, not all teachers managed to have all students write and rewrite all texts. This may have contributed to class differences in the effectiveness of the intervention. In addition, this observation highlights the importance of teacher support in having conversations about own texts. Finally, due the timing of the implementation a relatively large percentage of data (e.g., 25%) was missing on the far transfer posttest. Little's MCAR test revealed that this data was missing at random. Besides, multilevel models are well equipped to deal with missing data (Snijders & Bosker, 2012).

Finally, follow-up research is needed on how students discussed and revised their texts during the lessons. While the results show that students' writing quality improved over time, it is still unclear how this can be explained by the steps of writing, talking, and rewriting. This requires a more in-depth analysis of what aspects are specifically discussed in the peer conversations and how writers use these reader responses for revision. In the interviews with teachers after the intervention study they revealed that students did become more aware of the importance of having another person read and discuss their text, as they also applied this to new writing assignments independent of the intervention. This is hopeful since audience awareness is an important aspect of the writing process (Flower & Hayes, 1981; Graham, 2018). However, the teachers also noted that the final step of revision was difficult for students. They observed during the lessons that students by no means used all the feedback to revise their text. This observation is consistent with previous research by Bogaerds-Hazenberg et al. (2017), showing that the way feedback is formulated affects the use of it for revision. Future research should reveal how and

when students use peer feedback for revision and what additional support, feedback, and instructions from teachers they need to do so more effectively.

Implications for educational practice

Based on the results of this intervention study, we can make some concrete recommendations for educational practice. First, the results confirm the effectiveness of dialogic writing in promoting students' writing performance. By reading and talking about their own texts, students write texts that are more persuasive to a reader. Teachers are therefore advised to encourage students to have small-group conversations after a first draft of the text using the conversation chart and then use this feedback to revise their text. Teachers have indicated that this structure of writing—talking—rewriting is easy to implement in the classroom. Also, over time, exchanging texts and discussing them becomes automatic for students. In this way, it becomes easier for students to see the connection between the different language skills of writing, reading, and speaking and to integrate them with each other in a meaningful way.

Second, the principles of dialogic writing can also be used for other text genres such as for writing poetry or an informative text. When selecting a writing assignment, teachers are advised to make sure that it is a meaningful task, for which good communication is essential. This makes talking about the texts easier.

Furthermore, several teachers indicated that good conversations do not just happen and that a basic level of dialogic conversation in the classroom is needed for these lessons to be successful. It is therefore recommended that sufficient attention is paid to the ground rules of dialogic classroom talk prior to the lessons, see Appendix 2. Preferably, these ground rules should be established together with the students and repeated regularly as students begin to engage in conversations. It is also important that teachers themselves set a good example of dialogic conversations about texts. Therefore, teachers are advised to ask open and challenging questions about the texts and encourage students to complement each other and ask follow-up questions, rather than answering questions themselves or giving directive feedback on the texts.

Taken together, a dialogic writing intervention seems to be a promising approach to teach students how to talk about their texts and to write and rewrite texts that are more persuasive to readers. With this study, we not only sought to optimize interventions for learning to write and engage in conversations, but also to develop more generic knowledge about how to successfully integrate different language domains into classroom practice. In doing so, we hope to have contributed to an effective integration of the various language domains.

Appendix 1

Conversation chart for students.

Dialogisch schrijven



SPARRING about each other's texts

SPARRING... how do you do it?

In groups of three you will discuss the texts you have written yourselves. In each round you will discuss one text, so after three rounds you will have talked about all three texts. During the discussion you will take notes, so that after the discussion you can improve your own text.

Step 1: read

Read the first text in silence. Ask yourself, "*What is the writer trying to make clear?*" Write down what you like and what could be better.

Step 2: the writer begins the conversation and asks:

Do you think my text is clear? Why?
And what could be better?

Is my tekst convincing? Why or why not?

Can I change something to make you understand my tekst better? Do you have any tips?

Step 3: respond to each other

- Can you tell something more?
- I don't understand it yet, can you explain it in another way?
- What do you mean?
- So you're saying...?
- Do you have another idea?
- Do you agree with it? Why or why not?

Step 4: the writer summarizes

Tell briefly how you will improve your text.

Appendix 2

Basic rules for dialogic classroom talk.

Dialogisch schrijven



Basic rules for classroom conversations

Good conversations don't just happen; this is a skill that students need to work on together. It helps to set some ground rules together for having conversations in class. Below are the six most important ground rules for having good conversations. Involve students in setting these basic rules so they have a good understanding of what the rules stand for and how to follow them. Repeat the rules regularly and make sure the rules are always clearly visible to students during group discussions.

1. Each student is allowed to contribute their own ideas/opinions
2. We listen to each other and respect each other's input
3. We give arguments for our ideas/opinions
4. We ask questions and inquire about the other's ideas/opinions
5. We respond to each other and give feedback
6. We really try to understand each other

Funding This research was supported by grant 40.5.20500.176 from the Netherlands Organization for Scientific Research (NWO). We have no conflicts of interest to disclose. We thank all the teachers and teacher trainers who made this research possible.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Al-Adeimi, S., & O'Connor, C. (2021). Exploring the relationship between dialogic teacher talk and students' persuasive writing. *Learning and Instruction, 71*, 101388. <https://doi.org/10.1016/j.learninstruc.2020.101388>
- Bereiter, C., & Scardamalia, M. (1987). *The Psychology of Written Composition*. Erlbaum.
- Bodewitz, L. (2020). *In gesprek met de lezer: Een exploratief onderzoek naar interactie met de lezer voorafgaand aan het reviseren* [Unpublished bachelor thesis]. Vrije Universiteit Amsterdam.
- Bogaards-Hazenberg, S., Bouwer, R., Evers-Vermeul, J., & van den Bergh, H. (2017). Daar maak ik geen punt van! feedback en tekstrevisie op de basisschool [I don't make an issue of that! Feedback and text revision in elementary school]. *Levende Talen Tijdschrift, 18*(2), 21–29.
- Bouwer, R., & Koster, M. (2016). *Bringing Writing Research into the Classroom: The Effectiveness of Tekster, A Newly Developed Writing Program for Elementary Students* [Unpublished doctoral dissertation]. Utrecht University.
- Bouwer, R., Van Steendam, E., & Lesterhuis, M. (2023). Assessing writing performance: Guidelines for the validation of writing assessment in intervention studies. In F. De Smedt, R. Bouwer, T. Limpo & S. Graham (Guest Eds.), *Conceptualizing, Designing, Implementing, and Evaluating Writing Interventions*. Brill: Studies in Writing Series.
- Bouwer, R., Koster, M., & Van den Bergh, H. (2023). Benchmark rating procedure, best of both worlds? Comparing procedures to rate text quality in a reliable and valid manner. *Assessment in Education: Principles, Policy & Practice*. <https://doi.org/10.1080/0969594X.2023.2241656>
- Bouwer, R., Béguin, A., Sanders, T., & Van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing, 32*(1), 83–100. <https://doi.org/10.1177/0265532214542994>
- Bouwer, R., Koster, M., & van den Bergh, H. (2018). Effects of a strategy-focused instructional program on the writing of upper elementary students in the Netherlands. *Journal of Educational Psychology, 110*(1), 58–71. <https://doi.org/10.1037/edu0000206>
- Casado-Ledesma, L., Cuevas, I., Van den Bergh, H., Rijlaarsdam, G., Mateos, M., Granado-Peinado, M., & Martín, E. (2021). Teaching argumentative synthesis writing through deliberative dialogues: Instructional practices in secondary education. *Instructional Science, 49*(4), 515–559. <https://doi.org/10.1007/s11251-021-09548-3>
- Chanquoy, L. (2001). How to make it easier for children to revise their writing: A study of text revision from 3rd to 5th grades. *British Journal of Educational Psychology, 71*(1), 15–41. <https://doi.org/10.1348/000709901158370>
- Dobber, M. (2018). *Leerkrachten en leerlingen duiken samen de geschiedenis in* [Teachers and students dive into history together]. Didactief. <https://didactiefonline.nl/artikel/leerkrachten-en-leerlingen-duiken-samen-de-geschiedenis-in>
- Fisher, R., Jones, S., Larkin, S., & Myhill, D. (2010). *Using Talk to Support Writing*. Sage.
- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365–387. <https://doi.org/10.2307/356600>
- Graham, S. (2018). A revised writer(s)-within-community model of writing. *Educational Psychologist, 53*(4), 258–279. <https://doi.org/10.1080/00461520.2018.1481406>
- Graham, S., McKeown, D., Kihara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology, 104*(6), 396–407. <https://doi.org/10.1037/a0029185>

- Harris, K. R., Camping, A., & McKeown, D. (2023). A review of research on professional development for multicomponent strategy-focused writing instruction: Knowledge gained and challenges remaining. In F. De Smedt, R. Bouwer, T. Limpo & S. Graham (Guest Eds.), *Conceptualizing, Designing, Implementing, and Evaluating Writing Interventions*. Brill: Studies in Writing Series.
- Hennessy, S., Rojas-Drummond, S., Higham, R., Márquez, A. M., Maine, F., Ríos, R. M., Barrera, & M. J. (2016). Developing a coding scheme for analysing classroom dialogue across educational contexts. *Learning, Culture and Social Interaction*, 9, 16–44. <https://doi.org/10.1016/j.lcsi.2015.12.001>
- Herder, A., Berenst, J., de Glopper, K., & Koole, T. (2018). Reflective practices in collaborative writing of primary school students. *International Journal of Educational Research*, 90, 160–174. <https://doi.org/10.1016/j.ijer.2018.06.004>
- Hox, J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel Analysis: Techniques and Applications*. Routledge.
- Inspectorate of Education (2019). *Peil.Mondelinge taalvaardigheid einde basisonderwijs 2016–2017* [Assess.Oral language skills end of (special) primary education 2016–2017]. Utrecht: Inspectie van het Onderwijs.
- Inspectorate of Education (2021). *Peil. Schrijfvaardigheid einde (speciaal) basisonderwijs 2018–2019* [Assess.Writing end of (special) primary education 2018–2019]. Utrecht: Inspectie van het Onderwijs.
- Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research*, 1(1), 1–26. <https://doi.org/10.17239/jowr-2008.01.01.1>
- Kooijman, J. (2022). Jij en ik en al het moois om ons heen [You and me and all the beauty around us]. *Zone*, 21(4).
- Koster, M., Bouwer, R., & van den Bergh, H. (2017). Professional development of teachers in the implementation of a strategy-focused writing intervention program for elementary students. *Contemporary Educational Psychology*, 49, 1–20. <https://doi.org/10.1016/j.cedpsych.2016.10.002>
- Koster, M., Tribushinina, E., De Jong, P., & van den Bergh, H. (2015). Teaching children to write: A meta-analysis of writing intervention research. *Journal of Writing Research*, 7(2), 249–274. <https://doi.org/10.17239/jowr-2015.07.02.2>
- Myhill, D. A., & Newman, R. (2019). Writing talk: Developing metalinguistic understanding through dialogic teaching. In N. Mercer, R. Wegerif, & L. Mercer (Eds.), *Routledge International Handbook of Research on Dialogic Education* (pp. 360–372). Routledge.
- National Center for Education Statistics. (2012). *The Nation's Report Card: Writing 2011 (NCES 2012–470)*. Washington, D.C: Institute of Education Sciences, U.S Department of Education.
- Resnick, L., Asterhan, C., & Clarke, S. (2015). *Socializing Intelligence through Academic Talk and Dialogue*. AERA.
- Rijlaarsdam, G., Braaksma, M., Couzijn, M., Janssen, T., Kieft, M., Raedts, M., ... & Van den Bergh, H. (2009). The role of readers in writing development: Writing students bringing their texts to the test. *The SAGE Handbook of Writing Development*, 436–452.
- Smith, S. W., Daunic, A. P., & Taylor, G. G. (2007). Treatment fidelity in applied educational research: Expanding the adoption and application of measures to ensure evidence-based practice. *Education & Treatment of Children*, 30(4), 121–134. <https://doi.org/10.1353/etc.2007.0033>
- Bosker, R., & Snijders, T. A. (2011). Multilevel analysis: An introduction to basic and advanced multilevel modeling. *Multilevel Analysis*, 1–368. Sage
- Sperling, M. (1996). Revisiting the writing-speaking connection: Challenges for research on writing and writing instruction. *Review of Educational Research*, 66(1), 53–86. <https://doi.org/10.2307/1170726>
- Van den Bergh, H., & Eiting, M. H. (1989). A method of estimating rater reliability. *Journal of Educational Measurement*, 26(1), 29–40.
- Van der Veen, C., de Mey, L., van Kruistum, C., & van Oers, B. (2017). The effect of productive classroom talk and metacommunication on young children's oral communicative competence and subject matter knowledge: An intervention study in early childhood education. *Learning and Instruction*, 48, 14–22. <https://doi.org/10.1016/j.learninstruc.2016.06.001>
- Van der Veen, C., & van Oers, B. (2017). Advances in research on classroom dialogue: Learning outcomes and assessments. *Learning and Instruction*, 48, 1–4. <https://doi.org/10.1016/j.learninstruc.2017.04.002>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.