Oral sentence generation training to improve fifth and 10th graders' writing

Barbara Arfé[1] · Federica Festa[3] · Lucia Ronconi[2] · Gaia Spicciarelli[1]

Abstract

Text generation—the mental translation of ideas into language at word, sentence, and discourse levels—involves oral language abilities. However, oral language skills are rarely a target of writing interventions. We ran an intervention to improve fifth and 10th graders' written production through the development of oral sentence generation (grammatical and syntactic) skills. One hundred and fifteen students—68 fifth graders (four classrooms) and 47 tenth graders (four classrooms)—participated in a stepped-wedge cluster-randomized controlled trial. Two fifth-grade classrooms ($n = 35$) and two 10th-grade classrooms ($n = 20$) received nine 90-min sessions (3 weeks, three sessions a week) of oral language intervention immediately after the pretest (experimental groups); the two other fifth- ($n = 33$) and 10th-grade classrooms ($n = 27$) received business-as-usual writing instruction and received a delayed oral language intervention after the posttest (waiting list group). The intervention consisted of team-based games to improve oral sentence generation and sentence reformulation skills. We assessed written sentence generation, written sentence reformulation, written text quality (macrostructure and language), and text writing fluency before (pretest) and after (posttest) the intervention and 5 weeks after the intervention (follow-up). The results showed that training on oral sentence generation skills can lead to significant gains in both sentence generation and sentence reformulation skills and text macrostructural quality. Improvement at the sentence level was, however, significant only for the younger writers (fifth graders).

Keywords Sentence generation · Writing fluency · Writing quality · Oral language intervention · Struggling writers

✉ Barbara Arfé
barbara.arfe@unipd.it

1   Department of Developmental and Social Psychology (DPSS), University of Padova, Via Venezia 8, 35130 Padua, Italy

2   University of Padova, Padua, Italy

3   IIS C.I. Giulio, Torino, Italy

Text generation—the process of encoding ideas into language for writing at word, sentence, and discourse levels—involves oral language, as writers use their oral vocabulary and grammatical knowledge to generate words and sentences for their texts (Abbott & Berninger, 1993; Babayigit & Stainthorp, 2010). Yet, oral language skills are rarely a target of writing interventions (Goldfeld et al., 2017; Spencer & Petersen, 2018). The lack of intervention studies targeting text generation skills through oral language training appears surprising when one considers that difficulties in learning to write and read are often caused by deficits in underlying oral language skills (Dockrell & Connelly, 2013; Dockrell, Lindsay, Connelly, & Mackie, 2007; Hulme & Snowling, 2014). Even when writing difficulties are not associated with clear language impairments, oral and written expression are strongly related (Berninger & Abbott, 2010; Mehta, Foorman, Branum-Martin, & Taylor, 2005).

In line with other recent initiatives (Goldfeld et al., 2017), the present study was aimed at filling the extant evidence gap in the effectiveness of classroom-level oral language interventions on students' writing. We did this by testing the effects of a focused oral-language (i.e., sentence generation) intervention on Italian fifth and 10th graders' text generation and written composition skills.

The role of oral language and text generation skills in writing development

Oral language abilities are not always an explicit component of developmental writing models (see, for example, Berninger, 2000; Berninger et al., 2002). Yet, these models assume that oral language underpins text generation, the core component of the developing writing process (Berninger et al., 2002; Kim, 2016; Kim & Schatschneider, 2017). Indeed, text generation is often operationalized in writing research as oral vocabulary, grammatical, and discourse skills (Abbott & Berninger, 1993; Kim & Schatschneider, 2017).

The not-so-simple view of writing describes writing development as the product of the development of three interacting processes: text generation, which involves oral language skills, transcription (spelling and handwriting), and executive function (including attentional control, planning, review, and self-regulation abilities; Berninger, 2000; Berninger et al., 2002). The limited transcription skills (spelling and handwriting) of beginning writers (Abbott, Berninger, & Fayol, 2010; Berninger, 2000; Berninger et al., 1992; Graham, Berninger, Abbott, Abbott, & Whitaker, 1997; Pinto, Tarchi, & Bigozzi, 2015) hinder the translation of their oral language knowledge (vocabulary, grammatical and oral discourse skills) into writing and thus the potential impact of their oral language on writing. However, as writers get older and transcription skills automatize, the proportion of variance accounted for by transcription processes in writing drops (Berninger, 1999) and individual differences in text generation (and oral language skills) become more discriminant and start to influence writing more directly (Berninger, Nagy, & Beers, 2011; McCutchen, Covill, Hoyne, & Mildes, 1994). This process may require up to 4 years of writing instruction for English-speaking children (Berninger et al., 2011; Juel, 1988; McCutchen et al., 1994), but only 2 or 3 years for students writing in shallow

orthographies, such as Italian (Arfé, Dockrell, & De Bernardi, 2016; Arfé, Cona, & Merella, 2018; Babayigit & Stainthorp, 2010).

Recently, research conducted with English-speaking beginning writers has also suggested that developmental models of writing might have underestimated the role of oral language skills in early written composition (e.g., Kent, Wanzek, Petscher, Al Otaiba, & Kim, 2014; Kim, Al Otaiba, Wanzek, & Gatlin, 2015). This has led to the development of the direct and indirect effects model of writing (DIEW, Kim, 2016; Kim & Schatschneider, 2017). In contrast to the not-so-simple view of writing, the DIEW model of writing (Kim & Schatschneider, 2017) considers oral language skills to be a direct influence on written composition from the beginning of formal writing instruction, proposing that not only transcription skills but also oral discourse-level skills (a component of text generation) have a strong and direct effect on written production beginning in Grade 1. Similarly to the not-so-simple view of writing, the DIEW model of writing has been validated by empirical evidence of early effects of oral discourse-level language skills on writing (Kim, 2019; Kim & Schatschneider, 2017), which has been observed in older writers as well (fourth graders; Kim, 2019).

Other studies have explored more closely the interrelationship between oral and written language abilities across ages and grade levels (Berninger & Abbott, 2010; Mehta et al., 2005), showing that oral and written language are independent yet strongly interrelated constructs. For example, Mehta et al. (2005) found that, in elementary school children (third graders), the correlation between the two factors was .70 at the student level and 1.00 at the classroom level, where the two competences were indistinguishable. Berninger and Abbott (2010) found strong and significant correlations between oral language factors (listening comprehension and oral expression) and writing in Grade 1 (.65 correlation between writing and listening comprehension and .62 correlation between writing and oral expression), Grade 3 (.70 and .68), and Grade 5 (.67 and .66). These findings confirm that oral and written expression, though not identical, draw on common processes.

In line with this research, a recent meta-analysis (Graham, Hebert, Fishman, Ray, & Rouse, 2020) showed a consistent and significant association between oral and written language problems from the preschool years to adolescence (Grade 12). According to the authors, three theoretical explanations can be given for the link between oral language and writing difficulties. The first, the *shared knowledge* hypothesis, is that oral and written production share knowledge representations (phonological, morphological, and syntactic knowledge) and processes (e.g., translating ideas into language units, monitoring language production), and that problems in constructing these representations or in these processes thus affect both production modalities. The second explanation, the *rhetorical relations* hypothesis, is that children need to develop similar pragmatic and rhetorical skills (e.g., sense of audience and the ability to choose appropriate rhetorical devices) in the two production modalities. Children's problems in developing these rhetorical and pragmatic skills in oral production can influence writing development as well. A third explanation posits that the same general *learning deficit* or basic dysfunctions (e.g., a processing capacity limitation or deficits in procedural memory) can cause oral and written language difficulties (the *learning deficit* hypothesis). In all cases, empowering skills

and processes in one modality (oral language) should translate into benefits to the other modality (written language).

The increasing awareness of the role played by oral language skills in writing has recently led researchers to test the benefits of introducing oral language interventions into writing or literacy instructional programs (Goldfeld et al., 2017; Spencer & Petersen, 2018). Yet the research in this field is still limited, and these intervention studies have been focused on young writers only (first to third graders).

Although the foundational oral language components of text generation are partly developed before children learn to write (Kent et al., 2014; Pinto et al., 2015), they continue to develop concurrently with written language during upper elementary school years and beyond (Alamargot et al., 2015; Chanquoy & Negro, 1996; Hebert, Bohaty, Nelson, & Roehling, 2018, Hebert, Kearns, Hayes, Bazis, & Cooper, 2018; Jones, Myhill, & Bailey, 2013; Ravid & Tolchinsky, 2002; Wijekumar et al., 2019). For example, Ravid and Berman (2006) demonstrated that children's oral and written productions in English and Hebrew showed parallel developmental patterns from 9–10 years (Grade 4) to 16–17 years of age (high school), with a turning point in the development of oral and written language abilities in adolescence between ages 12–13 and 16–17. Thus, training in oral language skills also could be beneficial for these older writers.

The efficacy of text generation interventions and interventions focused on sentence-level skills

Limited evidence exists in general on the efficacy of instructional programs that target text generation directly and primarily (McMaster, Kunkel, Shin, Jung, & Lembke, 2018). In a best-evidence review, McMaster et al. (2018) showed that 17 out of 25 studies addressing Grades 1 to 3 targeted text generation in multicomponent interventions; however, most of them did not examine the specific contribution of training text generation processes. Other meta-analyses that examined the effectiveness of writing interventions with elementary school children (Graham, McKeown, Kiuhara, & Harris, 2012; Graham & Perin, 2007) or older writers (Graham & Perin, 2007) analyzed four types of instructional programs focused primarily on text generation—explicit teaching of grammar, sentence combining, teaching text structure, and emulating good texts—and reported only small to moderate efficacy (effect sizes from 0.25 to 0.59) for those programs. Of the instructional strategies assessed, sentence combining (i.e., teaching students to construct compound, complex or sophisticated sentences) and text structure interventions (i.e., providing explicit knowledge of the text structure of specific text types) led to the largest effect sizes (0.50 and 0.59, respectively). Grammar instruction was the sole treatment that appeared to have a negligible or even negative effect on students' writing (Graham et al., 2012).

Independently of the limited efficacy of grammar instruction, however, the fluency with which writers generate sentences (i.e., grammatical and syntactic structures) orally or in writing appears to be a crucial element of the writing performance of young writers (e.g., Arfé et al., 2016). In fact, in upper elementary school,

measures of writing fluency are most effective in discriminating between good and struggling writers (Dockrell, Connelly, & Arfé, 2019; McCutchen et al., 1994).

Text writing fluency is an important feature of skilled writing, reflecting the facility with which writers generate and produce connected texts (Alves & Limpo, 2015; Dockrell et al., 2019; Kim, Gatlin, Al Otaiba, & Wanzek, 2018; Limpo & Alves, 2018). Writing research has operationalized writing fluency in several different ways: (a) as the number of words produced in a text per minute (i.e., the speed of transcription and text generation; Limpo & Alves, 2018); (b) as the number of words produced in a writing burst (i.e., between writing pauses, a second speed measure; Alves & Limpo, 2015; Dockrell et al., 2019); or (c) as the number of correctly connected words produced within a time limit (i.e., the accuracy and speed of transcription and text generation; Kim et al., 2018). These measures tap different aspects of writing fluency. The first two measures (a and b) tap the automaticity of transcription and text generation. The third measure (c) taps the efficiency with which children execute the transcription and text generation task (i.e., how well and rapidly they perform the task). In both cases, in beginning writers, transcription skills mainly constrain text fluency. However, as noted earlier, later in writing development—from Grade 3—the contribution of oral language (text generation) skills to text writing fluency increases. As words in a text connect by means of grammatical and syntactic rules, sentence fluency (fluency in generating grammatical and syntactic structures) becomes an important index of writing ability at this point in writing development (Arfé et al., 2016; Dockrell et al., 2019; McCutchen et al., 1994). In this study we operationalized text writing fluency and sentence fluency as a writer's speed of producing connected text (words in a text or complete sentences).

A writer's ability to fluently generate and handle complex sentence structures can be critical, especially in romance languages, such as Italian (Arfé et al., 2016), where a very shallow orthography may reduce the demands of spelling, but the morphological and grammatical complexity of the language increases the demands of text generation (Arfé et al., 2016). Thus, sentence generation skills not only play a significant role in the quality and complexity of the texts produced by young Italian writers (Arfé et al., 2016), but also continue to account for variance in writing fluency (i.e., speed) even during high school years (ages 16–18; Danzak & Arfé, 2016). The development of the cognitive and neural underpinnings of syntactic processing is indeed prolonged, continuing through adolescence (Schneider, Abel, Ogiela, Middleton, & Maguire, 2016), and the ability to organize information in complex sentences by means of a variety of subordinate conjunctions represents one of the areas of greatest weakness for adolescents with language learning needs (Gamez, Lesaux, & Rizzo, 2016).

Training oral language abilities at sentence and intersentential levels may be thus an effective strategy to improve writing skills, especially (although not only) in a morpho-syntactically complex language. Past intervention studies have shown that sentence-level interventions focused on sentence construction can be effective for writers of different ages (from upper elementary school to high school), and the positive effects can transfer to extended composition (Datchuk & Kubina, 2013; Limpo & Alves, 2013; Myhill, Jones, Lines, & Watson, 2012; Saddler & Graham, 2005). However, the focus of these interventions was on written sentence construction. To

our knowledge, no research has attempted to test oral sentence generation skill training. Training oral sentence construction and fluency can be beneficial for writers, allowing for focus on the linguistic processes of text generation without the additional demands of transcription, which may be a constraint for translation processes even during middle school years (Grades 7–8; Limpo, Alves, & Connelly, 2017) and up to adulthood (Olive & Kellogg, 2002). Such interventions could be particularly effective for those writers who show weak foundational oral language skills because of language disabilities (Dockrell, Lindsay, & Connelly, 2009) or socioeconomic or linguistic disadvantages (Hoff, 2013). Empowering the oral language skills of these writers could have beneficial effects on their writing outcomes.

In the present study, we aimed to explore the effectiveness of a nine-session, classroom-level oral language intervention focused on the oral sentence generation skills of students in Grades 5 and 10. As noted earlier, students' linguistic skills in oral and written language abilities continue to develop from 9–10 years of age (Grade 4) to 16–17 years (high school; Ravid & Berman, 2006). Grades 5 and 10 represent two important turning points in writing development. Grade 5 is the last year of elementary school in Italy. At this stage, written production tasks become more complex and varied to prepare children for middle school compositional tasks. Fluency in text generation represents a prerequisite and developmental constraint for engaging in these more complex tasks (Dockrell et al., 2019). Grade 10 is the last grade in which grammatical skills are taught in Italian school and represents a consolidation stage of higher level syntactic skills and metalinguistic abilities that are necessary for rhetorically complex writing tasks, such as persuasive writing (Brimo & Hall-Mills, 2019), as well as academic writing in general (Silliman, 2014). We addressed the intervention to classrooms in which teachers reported significant writing needs among students and targeted oral sentence generation and reformulation skills, progressing through increasing levels of difficulty, from simple to complex sentence construction up to the use of inter-sentential links at the discourse level. We examined the effectiveness of the training in improving written sentence fluency (close transfer) and writing quality and fluency (far transfer). Three research questions guided this study:

1. Would an oral language intervention focused on sentence and inter-sentential construction skills yield significant gains at the sentence and text levels in fifth- and 10th-graders' writing?
2. Will potential gains differ depending on students' grade level or level of writing development? Related to this, will the intervention affect text generation at different levels (i.e., sentence and text level) in fifth-grade and 10th-grade students' writing?
3. Will students retain writing gains post-intervention?

We made the following hypotheses:

- Given that the cognitive underpinnings of syntactic skills continue to develop until adolescence (Schneider et al., 2016), we expected that training oral sen-

tence generation skills would lead to significant improvements in students' written sentence generation fluency and written sentence reformulation skills at both grade levels (Grades 5 and 10). However, we expected greater improvements at the sentence level for the younger writers (fifth graders), whose sentence-level text generation skills were assumed to be less mature.

- Given the role played by oral language and sentence generation fluency in students' writing (Dockrell et al., 2019), students' gains following the training should transfer to text production, impacting the fluency (text generation speed) and quality of the students' written compositions.

- Students should maintain these positive effects at a 5-week follow-up from the end of the intervention.

The study

Participants

We enrolled two groups of participants in the study, one including four classrooms of young writers attending the last year (Grade 5) of an Italian public primary school in northwest Italy, and one including four classrooms of older writers attending the second year of an Italian public high school (Grade 10) in the same region. Inclusion criteria for the study, at the classroom level, were: (a) a classroom reported by the language teacher to present writing needs; and (b) teachers' commitment to participating in the full (nine sessions) classroom-level intervention. Inclusion criteria at the student level included: (a) residing in Italy for at least 3 years; (b) no reported sensory, motor, or intellectual disabilities; and (c) parents' written informed consent to the study. Currently, there are no specific formulae for the computation of power and sample size in cohort stepped-wedge trials involving repeated measures. The required sample size, computed for a factorial repeated-measures design, is N = 128, with power set at .80 and effect size at .25, $p < .05$, and N = 171, with power set at .90. Based on these calculations, we initially recruited 167 participants (81 fifth graders and 86 10th graders) to take part in the study. Of these, only 115 students participated in all the assessment phases of the study (pretest, posttest, and follow-up) and in a sufficient number of training sessions (at least 7 out of 9) to be included in the final study sample. Thus, attrition was 31%. The dropout rate was higher among high school students (n = 39 participants; 45%), probably due to the relatively greater socioeconomic disadvantage of these students in comparison to the elementary school students ($\chi^2 = 109.33$, $df = 2$, $p < .001$). Students from lower socioeconomic backgrounds typically show higher rates of school absenteeism (Klein, Sosu, & Dare, 2020).

The final study sample comprised of 68 primary school children (37 girls, 54%, mean age = 9.92) from four fifth grade classrooms, and 47 high school students (37 girls, 78%, mean age = 15.54) from four, 10th grade classrooms. Table 1 reports the participant characteristics. Four fifth graders and four 10th graders of this sample were diagnosed with learning disabilities: three fifth graders and one 10th grader presented a mixed learning disorder, including deficits in reading,

Table 1 Pretest (T1) differences between the experimental and wait-list groups: means (standard deviation)

Variables	Grade 5			Grade 10		
	Experimental (n=35; 19 girls)	Wait list (n=33; 18 girls)	p	Experimental (n=20; 17 girls)	Wait list (n=27; 20 girls)	p
Mean age (years)	9.81 (0.54)	10.06 (0.41)	.032	14.47 (0.85)	15.59 (0.98)	.659
Girls (%)	54%	54%	.983	85%	74%	.366
With LDs (%)	1	3	.275	2	2	.753
Fewer than 6 years in Italy (n)	2	1	.590	1	3	.458
ZSpelling	.02 (.51)	−.03 (1.35)	.835	0.31 (0.92)	−0.23 (1.01)	.067
Sentence-level skills						
Sent_Gen_Acc	8.96 (3.63)	9.09 (3.92)	.890	12.29 (3.62)	11.31 (5.05)	.467
Sent_Gen_Fluency	5.49 (1.82)	6.26 (2.82)	.182	6.45 (1.64)	5.56 (2.19)	.132
Sent_Reformulation	2.16 (1.04)	2.64 (1.18)	.085	4.30 (0.81)	3.80 (1.24)	.121
Text-level skills						
Macrostructure	3.69 (1.35)	4.03 (1.40)	.305	4.35 (1.04)	4.63 (1.04)	.368
Language	2.09 (0.82)	1.76 (0.71)	.082	2.10 (0.64)	2.37 (0.74)	.198
Text fluency (words in 15 min)	89.23 (31.51)	119.97 (50.50)	.003	182.60 (78.03)	128.67 (35.51)	.003

Fewer than 6 years in Italy=Students having lived in Italy for fewer than 6 years; with LDs=students with learning disabilities; ZSpelling=spelling Zscores; Sent_Gen_Acc=mean sentence generation accuracy; Sent_Gen_Fluency=mean sentence generation fluency; Sent_Reformulation=mean sentence reformulation score; Macrostructure=mean text macrostructure score; Language=mean text language score; Text fluency=mean fluency in text production (words per 15 min)

writing, and math; and three 10th graders and one-fifth grader had other kinds of learning disabilities (including visuospatial or attentional disorders). Twenty-seven participants (23% of the sample, seven fifth graders and twenty 10th graders) were from other countries of origin (Romania, Albania, Senegal, China, Brazil, Morocco, and Russia) and learned Italian (their second language, L2) at school. Only seven of these students (three-fifth graders and four 10th graders) were in Italy for less than 6 years (range 3–5 years) at the time of the study. They were fluent enough in Italian to take part in the study activities.

We did not obtain the school's permission to collect students' socioeconomic (SES) data. However, the two school districts were respectively in a medium (for fifth graders school) and medium to low (for the 10th graders' school) SES areas. The participants who dropped from the study did not differ significantly from the other study participants on any of the demographic variables considered (i.e., age, gender, SES area, learning disabilities, or years in Italy). Little's MCAR (missed completely at random) test run on pretest scores showed that the students with missing data were not significantly different from the participants

without missing data: $\chi^2 = 13.15$, $p = .28$ (fifth graders) and $\chi^2 = 20.24$, $p = .09$ (10th graders).

For each grade level (5th or 10th), we randomly assigned classrooms to an experimental (two fifth grade classrooms, n = 35; two 10th grade classrooms, n = 20) or a wait-list condition (two fifth grade classrooms, n = 33; two 10th grade classrooms, n = 27). The experimental and wait-list groups did not differ significantly on gender distribution, students' years in Italy, and prevalence of learning disabilities (see Table 1). The only significant difference was the age between the fifth grade experimental and wait-list group $t(66) = -2.19$, $p < .05$: the children in the wait-list group were slightly older than the children in the experimental group were (see Table 1).

All parents provided written informed consent for their children to participate in the study. In addition, we requested students' oral consent to the study before each assessment session.

Procedure

We ran a stepped-wedge cluster-randomized controlled trial (Campbell et al., 2019) with groups (classrooms) randomly assigned to an experimental or wait-list condition and both groups (the experimental and wait-list) receiving the intervention at different times. Figure 1 displays the study timeline. This experimental design allowed all students to receive the oral language intervention, while also controlling for intervention effects. Participants' sentence-level and text-level writing skills were assessed three times over the study period (see Fig. 1): pretest (T1), posttest (T2, after 3 weeks of intervention), and follow-up (T3, 5 weeks after the posttest).

The Grade 5 and Grade 10 experimental groups received the intervention immediately after the pretest, between the end of November and December. During this time span, the wait-list groups received business-as-usual (BAU) writing instruction. After the 3-week experimental/BAU intervention, at the end of December, we reassessed the fifth grade and 10th-grade students' sentence-level and text-level writing skills (T2, posttest). Next, the wait-list group received the experimental training for an equivalent duration (3 weeks in January), and the experimental group followed standard (BAU) writing instruction in their classrooms. At the end of January (T3), after this second 3-week intervention, we reassessed the students' sentence-level and text-level writing skills. For the experimental group, this second posttest, 5 weeks after the first, consisted of a follow-up assessment.
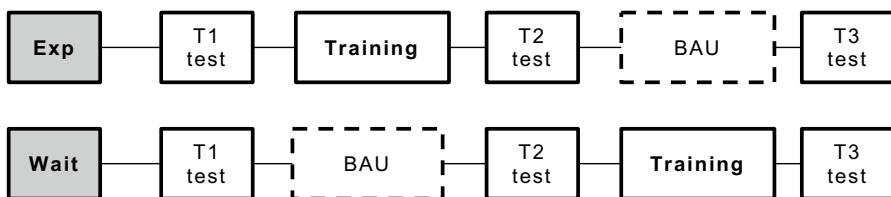

Fig. 1 Study design

Pretest and posttest assessments

We assessed the students' sentence-level and text-level writing skills at the pretest (T1), posttest (T2), and at the follow-up (T3) by written sentence generation and sentence reformulation tasks (Arfé et al., 2016), and written composition tasks (fantasy stories). We used parallel versions of the written sentence generation and reformulation tasks and similar text production tasks to retest sentence-level and text-level writing skills over time. Students completed all tasks in one classroom session of approximately 1 h. In addition to these tasks, all participants completed a standardized spelling task in the pretest session to assess their transcription skills.

Standardized spelling tasks

Different standardized spelling tests are used in Italy to assess spelling during elementary school and high school years. Therefore, we used two different spelling tasks for the two age groups of this study: the fifth and 10th graders.

Word Dictation (Grade 5) We used the word-spelling subtest of the Battery for the Assessment of Dyslexia and Dysorthographia (DDE-2; Sartori, Job, & Tressoldi, 2007) to assess the fifth graders' spelling abilities. Participants wrote, from dictation, 48 words varying in length (from two- to four-syllable words), frequency, and orthographic structure. We scored the number of misspelled words. The concurrent validity of this subtest, reported by prior studies (Arfe et al., 2016), is .82.

Text Dictation (Grade 10) For the 10th graders, we used a standardized text dictation task from the Neuropsychological Assessment Battery for Adolescence (BVN 12–18; Gugliotta, Bisiacchi, Cendron, Tressoldi, & Vio, 2009) to assess spelling skills. This test consists of writing a short text (152 words) under dictation within a 2-min time limit. We asked students to pay attention to the dictation and to write down the text as accurately and as rapidly they could. The number of correctly spelled words were scored. The manual does not provide reliability or validity values for this subtest, which researchers extensively use in clinical settings.

Sentence-level tasks We used two tasks that have shown good discriminant validity in the assessment of students' text generation skills (Arfé et al., 2016; Dockrell et al., 2019).

Sentence generation This task provides a sentence-level measure of students' text generation fluency, tapping their ability to generate ideas and to translate them in written sentences (syntactic and semantic structures; Arfé et al., 2016). Students received two word pairs (e.g., *acqua-ponte*/*water-bridge* and *cani- gatti*/*cats-dogs*) and were asked to generate as many different sentences as they could using both words in 5 min. We scored only sentences containing both words and allowed no changes to the word pair (e.g., changing word form from singular to plural). Participants practiced with an example word pair before starting the test. We recorded sentence fluency (number of different sentences generated) and accuracy (sentence generation accuracy scores). For accuracy, we adapted the scoring from Arfé et al. (2016) and Dockrell et al.'s (2019) studies, allowing for better discrimination among older students' sentence generation skills. Scoring criteria were:

- Each new sentence that was grammatically and semantically accurate earned a score of 2;
- Each sentence that was either semantically or grammatically incorrect earned a score of 1;
- Each sentence that was semantically and grammatically incorrect, did not include both target words, or included variations of the target word pair (e.g., plural instead of singular) earned a score of 0;
- Sentences that varied minimally from the previous ones (e.g., *The fireman jumped in the water from the bridge* and *The old lady jumped in the water from the bridge*) earned a score of 0.5.

We did not code errors in punctuation, capitalization, or misspellings.

Reliability The second author and a trained independent rater (a master's student), who was blind to the research groups and to the hypotheses of the study, independently scored the sentences. Interrater agreement computed on 100% of the sentences was high, 95%, and similar to that found by Dockrell et al. (2019): 94%. The two raters discussed and solved any disagreements.

Sentence reformulation

This was an additional, sentence-level measure of text generation fluency based on higher-order metalinguistic skills. The task assessed students' ability to reformulate sentences (i.e., find alternative words or grammatical structures to translate a given meaning; Arfé et al., 2016). Students were asked to reformulate two simple (one clause) and two complex (main plus subordinate clause) sentences. For each sentence, the student could generate up to three reformulations by using different words (i.e., synonyms or paraphrases) and/or grammatical/syntactic structures (e.g., translating sentences to passive voice). Differently from the sentence generation task, in the sentence reformulation task students were constrained by semantics; that is, the reformulated sentence needed to convey the same meaning as the given sentence. A time limit of 10 min was given for each trial. Before starting the test, the children practiced with two training items. The scoring was the same as that used in prior studies (Arfé et al., 2016):

- A score of 2 was awarded if the sentence reformulation was grammatically correct and maintained the meaning of the target sentence (e.g., *Alice wishes to play cards with Lucia* for *Alice wants to play cards with Lucia*).
- A score of 1 was given to reformulations that were grammatically correct but did not maintain the original meaning of the item (e.g., *Alice plays cards with Lucia* for *Alice wants to play cards with Lucia*).
- A score of 0 was given when the reformulation was incorrect both grammatically and semantically or the reformulated sentence was totally unrelated to the target (e.g., *Alice makes a cake with Lucia* for *Alice wants to play cards with Lucia*). Scores could range from 0 to 24 (2 [maximum score]×3 reformulations×4 sentences).

As with the sentence generation task, errors in punctuation, capitalization, or misspellings were not scored.

Reliability

Arfé et al. (2016), who tested younger writers (6- to 8-year-old children), reported an interrater reliability of 93%. In this study, interrater agreement between the second author and the independent rater was 86% on 100% of the sentence reformulations. All cases of disagreements were discussed between the two raters and solved.

Text-level tasks

Written composition Students were asked to write fantasy stories based on a topic title. Similarly to other studies (Dockrell et al., 2019), 15 min were allotted to complete the task. Three topic titles were given in the three (T1, T2, and T3) assessments: "An old man, feeling useless and seeking attention, decides to become a cat… [Continue]" (Topic 1, T1); "One morning, you wake up to discover you are an adult. Your parents have become children/adolescents. What happens next?" (Topic 2, T2); and "Today is the 1st of February of the year 3000. Retell your day" (Topic 3, T3). When 15 min had passed, students were asked to stop writing. If, at that moment, they were writing a sentence, they were allowed to complete it before stopping.

Texts were scored according to text quality and text generation fluency (number of words).

Text quality A scoring rubric was used to score text quality on a scale from 1 to 4. Two dimensions were scored, based on prior research (Arfé et al., 2016) and existing analytical writing scales (WOLD; Rust, 1996).

Macrostructure The macrostructural dimension included two dimensions of the WOLD: Ideas and development; and Organization, unity, and coherence (Rust, 1996). Text macrostructure was assessed considering the number of ideas produced and the quality of their logical organization (i.e., how well they were connected in a coherent text). Ideas were scored 1 when the text included few and poorly elaborated ideas and 4 when the text was rich with ideas, which were always articulated. Organization was scored 1 when the text presented a weak logical organization and several instances of incoherence and 4 for a very coherent and logically organized text. The two scores were summed for a macrostructure score (range 2–8).

Language The language dimension referred to the microstructure of the text, and included the Vocabulary, Sentence structure and variety, and Grammar and usage dimensions of the WOLD (Rust, 1996). The linguistic quality of the text was scored considering the writer's lexical choices, the grammatical and syntactic accuracy of the text produced, and the variety of syntactic structures used. Scores were higher (4) for texts with contextually appropriate lexical choices, syntactically accurate sentences, and a great variety of lexical and syntactic structures. A low score (1) was characterized by incorrect lexical choices, frequent repetitions, and several grammatical and syntactic errors.

Text fluency The number of words written in 15 min was used to score writing fluency (i.e., speed of text generation).

Reliability Interrater agreement computed on 100% of the texts, independently scored by the second author and the independent rater, was 81% for macrostructural quality, 82% for linguistic quality, and 100% for text fluency. Disagreements between the two raters were discussed and resolved. When reaching a consensus was difficult, the average score between the two raters was used.

The intervention

The intervention consisted of nine, 90-min language workshops conducted by the second author, with the assistance of the classroom teacher. The workshops aimed at stimulating the development of students' oral language abilities through group sentence generation games. Playful, teamwork activities (in groups of 4–5 members each) were designed to encourage students' interaction, engagement, and problem solving (Boscolo, Gelati, & Galvan, 2012). The linguistic games were aimed at (a) increasing students' awareness of sentence construction rules; and, (b) enhancing their sentence generation and linguistic fluency.

Linguistic games of different types (e.g., sentence re-construction, sentence expansion, sentence combining) alternated during the intervention to force students to generalize and flexibly use the linguistic abilities developed during the workshops. Workshop sessions and games were organized in levels of increasing difficulty. Games focused on simple sentence (verb plus complement) construction were introduced first, followed by games requiring the construction of longer and more complex sentences, involving the use of logical connectives and anaphors. Task constraints increased progressively to reflect the demands of text generation in writing, in which the production of new sentences is constrained by the text generated up to that moment. Finally, the students played games in which they had to transfer their sentence generation skills to argumentative or narrative tasks. Activities were discussed with the class at the end of each workshop, and informative and corrective feedback was provided to the teams.

Nine, 90-min workshops were administered to the experimental group (between T1 and T2) and to the wait-list group (between T2 and T3; see Fig. 1), distributed in three weekly workshops, during school hours, for a duration of 3 weeks (about 13 h of intervention in total).

Sessions were organized as follows.

Session 1. How do words and sentences affect meaning and message?

The goal of this introductory workshop was to frame the intervention and engage students in the workshop activities: A story (Grade 5) or an essay (Grade 10) around the theme of "language use" was read to the classes to prompt students' questions and classroom discussion. Questions related to how a text meaning might change depending on the words or sentences used, or questions around how specific lexical

or grammatical choices can strengthen a given message, were elicited. Then, the first game (broken sentences) was presented. In this session, students were not yet grouped in teams, but worked individually. This allowed the researcher to get to know the students and then assign them to mixed-ability teams in the second workshop session.

Broken sentences, basic

The students received a set of cards, each containing one of the elements of a broken sentence (a noun, a verb, a preposition, etc.). The task was to reconstruct the sentence in the best way possible, recombining the given sentence elements. Once the sentence was reconstructed, other students were asked to reduce its length as much as possible, up to its kernel, by deleting all unnecessary sentence elements one by one. Elements could be deleted if their deletion did not affect the grammaticality or semantic completeness of the sentence. The goal of this game was to invite students to reflect on the core role of verbs in sentences and the minimal elements necessary to convey a complete sentence meaning (see Tesnière, 2015).

Session 2

In this session, students were assigned to groups. Linguistic games were played in teams of 4–5 members each. The experimenter always introduced new games, showing the students how to perform the tasks. Two games were played in Session 2.

Sketches, basic

As the verb is the structural center of the clause, our sentence generation intervention started from verb valence. In linguistics, valence is the number of dependent elements (arguments or complements) that are necessary to complete a verb meaning (Tesnière, 2015). Like atoms combine in chemical compounds, words combine with other words in sentences. Valence is a measure of their combining capacity. Some verbs have greater valence than others, because they require more elements to convey a complete meaning. For example, the verb "to *give*" (Italian, *dare*) is trivalent, as it needs at least three complements (who *gives* what, to whom) to convey the action meaning, whereas the verb "to *rain" (*Italian*, piovere)* is monovalent, because it does not require complements (*piove*/it rains, conveys a complete meaning).

All teams received a box containing puzzle pieces representing different verbs. The task was to randomly pick up a puzzle piece (a verb) and represent the verb meaning in a "sentence sketch." The sketch should be "interpreted" by actors (team members), embodying all the sentence elements necessary to represent the verb meaning. Each team selected as many actors as necessary to convey the meaning of the target verb. For example, to represent the verb "to give," the team required a minimum of three "actor elements": an agent (who gives), an object (what), a receiver (to whom): *Marco gives the book to Luisa*. In turn, each team "interpreted"

the sentence for the classroom, explaining why a certain number of "actor elements" were necessary. The goal was to stimulate a reflection about verb valence.

Sketches, advanced

Similarly to the basic sketches game, teams picked up a verb puzzle piece from a verbs box and represented the action meant by the verb in a sketch. Differently from the basic sketches game, however, the task also involved developing the sentence further. Students were taught to use questions to decide how to expand the sentence meaning. For example, the action "*went*" could be represented in a simple sketch: "*He went to the park*". Then, questions such as "*with whom?*", "*why did he go there?*", and "*how did he get to the park?*" could help develop the sentence. One point was given to the team for each additional element that was correctly added to the basic sentence.

Session 3

In Session 3, students continued practicing sentence construction skills, but with increasing content and grammatical constraints. Generating sentences respecting specific constraints was considered an important target skill of the intervention, as text generation is always constrained by the preceding text, the words that have been used, and rhetorical constraints. Two games were played.

Sentence puzzles, basic

Each team received a box with 15 puzzle pieces corresponding to sentence elements. Each element had a different color depending on its grammatical category (e.g., light blue for articles, pink for verbs, violet for nouns, brown for prepositions, etc.). The task was to use as many puzzle pieces (sentence elements) as possible to generate a sentence. One point was awarded for each element (puzzle piece) added to the sentence. The longer the sentence, the higher the team's score. Grammatically incorrect sentences were scored 0.

Crazy sentences, basic (connectives)

For this game, teams were randomly given two picture cards depicting two content elements each (e.g., a *man* and a *penguin* on one card and a *king* and a *present* on the second) and were asked to formulate (within 20 s) a sentence that related the four elements. To combine all elements in a sentence, students had to use connectives. However, they were not allowed to use the additive connective, *and*. As noted in the introduction, students with language learning needs may lack the ability to organize information in complex sentences by means of conjunctions (Gamez et al., 2016). The students in this study tended to use "*and*" in place of more specific subordinate or coordinate connectives. By not allowing the use of "*and*", we pushed them to make use of more specific connectives in sentences. The team received one

point for each grammatically accurate sentence that did not include the connective *and*.

Session 4

In Session 4, an advanced version of the "sentence puzzles" game was proposed, in which the linguistic constraints of the game increased.

Sentence puzzles, advanced

This game was similar to the basic sentence puzzles game, but with the additional constraint that the team should use a specific verb to construct the sentence. All teams received a puzzle piece representing the target verb, plus 15 further puzzle pieces, corresponding to different sentence elements (nouns, articles, adjectives, etc.). Their task was to construct the longest sentence possible, starting from the sentence elements at hand. Like in session 2, verbs of different valence were used. Once the sentence was generated, the team could develop it further by rolling dice and picking up from different element boxes (e.g., nouns, adjectives, verbs boxes) as many sentence elements (nouns, articles, or other verbs) as the dice indicated. The team received as many points as the number of the correctly used sentence elements.

Session 5

As in Session 4, in Session 5, students only played one game. This session aimed at consolidating the sentence generation skills developed in Sessions 2 and 4 (mainly sentence expansion).

Syntactic goose game, basic

In this game, two teams competed in a challenge. The first player generated a sentence based on a verb card picked up from a verb card deck. Then, the teammates picked up a card each from another card deck (including Wh- question cards such as "*with whom?*" or preposition cards, like *of*). The task was to develop the sentence by answering the Wh- question (e.g., "*with whom?*") or using the target preposition (e.g., *of*) card. The number of cards correctly used to develop the sentence indicated how many steps along the track the team advanced. Then, a player on the opposing team picked up a verb card and generated a sentence containing the verb. The other members expanded the sentence by picking up a Wh- question or preposition card each.

In Sessions 6 and 7, the focus was on the use of inter-sentential links (connectives and pronominal references) and the construction of complex (main plus subordinate) sentences. As in sentence combining (Saddler, Behforooz, & Asaro, 2008; Saddler & Graham, 2005), the games were aimed at increasing students' awareness of the logical meaning of connectives and their ability to use them (as well as pronominal references) to integrate information from different sentences.

Session 6

In Session 6, fifth graders played the advanced broken sentences game, whereas 10th graders played the advanced crazy sentences game. The two games had the same goal but matched the different learning skills of the two age groups. We decided to use different games for the two age-groups to provide each group of students with an optimal challenge for their ability level.

Broken sentences, advanced

This game was the reverse of the basic broken sentences game. All teams were given scissors and a sheet of paper with a complex sentence printed on it: a main plus a subordinate clause, containing a temporal or causal connective. The task was to break the sentence into its basic elements and subsequently recombine them in different sentences, using the same temporal or causal connective used in the original sentence. One point was given for each element included in a grammatically and semantically correct reformulation.

Crazy sentences, advanced

This game was similar to the basic crazy sentences game. The teams had to generate sentences starting from two picture cards (four content elements) within 20 s. However, in the advanced version, the teams were also given a specific coordinating or subordinate connective (or a pronoun) to be used in the sentence. Additive connectives such as *and* were excluded. Teams were provided with a conjunctions table with connectives divided based on their logical function (causal, temporal, etc.) and were told that they could use the table to decide how to use the connective. The produced sentences were evaluated by the other teams and discussed with them.

Session 7

In this session, the students only played an advanced version of the syntactic goose game.

Syntactic goose game, advanced

This game was similar to the basic version, with the addition of conjunction and verb cards for sentence expansion. The expansion card deck contained Wh-questions (e.g., "*why?*") and conjunctions (e.g., *while*; instead of prepositions). The task was to develop a sentence generated based on a card verb, using Wh-questions (e.g., "*with whom*?") or conjunctions (e.g., *while*), with the additional challenge of adding a new verb to the sentence. Thus, the sentences constructed were more complex and included inter-sentential links.

In sessions 8 and 9 the goal was to stimulate students' transfer of their sentence combining skills in the context of discourse production. The focus was on the ability to use connectives to express logical relations between sentences and combine sentences in discourse.

Session 8

In Session 8, fifth graders played the story gaps game, whereas 10th graders played argumentative challenges. As in Session 6, the two games had the same goal but were adapted to match the different discourse skills of the two age groups.

Story gaps

In this game, knowledge of inter-sentential links was applied at the discourse level. A picture story sequence was given, with some gaps to fill. Students received a box of puzzle pieces corresponding to different conjunctions. The task was to tell the story, filling in the story gaps with appropriate logical connectives (e.g., *then*, *so*, *but*). One score was given for each correctly used connective.

Argumentative challenges

Taking turns, one member of each team confronted a member of another team to debate on a topic chosen by the group. The only constraint was to use some of the connectives from the conjunctions table used in the advanced crazy sentences game.

Session 9

In this final session, the groups worked together on a collaborative text construction. While the fifth-grade group worked on a narrative text, the 10th-grade group produced an argumentative text. At the beginning of the activity, each group received a set of connectives. To add a new idea to the text, each group had to generate at least one sentence using one of the connectives received.

Treatment fidelity

The first and second authors of the study jointly designed the training, and the second author administered the training and conducted all sessions with the assistance of the classroom teacher. The second author used a rubric to take note of any deviation from the intervention plan or problem with the planned training program. Deviations from the planned lessons steps were very rare and never exceeded 10% of the planned lesson. Instructional sections were timed and lasted 90 min each to ensure equivalence of treatment time spans across classrooms. To guarantee balanced participation among students, teamwork sessions were designed so that each team member had a turn being in charge of leading the task. Once a week, the first and second authors met online or physically to discuss any problems that arose during

the intervention. The only reported deviations from the planned intervention were some session shifts due to other scholastic activities, and all classrooms received the same sequence and length of intervention.

Data analyses

Gains in writing over the three time points (T1–T3) were assessed by computing writing gain scores (in sentence generation accuracy, sentence generation fluency, sentence reformulation, and written composition macrostructure, language, and fluency). Gain scores corresponded to the gains in students' performance from the pretest (T1) to the posttest (T2) and to the gains between the posttest (T2) and the follow-up (T3). T1–T2 gains were calculated by subtracting the students' scores at T1 from their scores at T2; T2–T3 gains were calculated by subtracting scores at T2 from those at T3. As the experimental group received the intervention between T1 and T2, the first gain score (T1–T2 gain) was considered a measure of students' learning following the intervention, whereas the second (T2–T3 gain) was considered a measure of maintenance of learning effects. For the wait-list group, the first gain score (T1–T2 gain) reflected only improvements due to practice, (i.e., test–retest effects), whereas the second (T2–T3 gain) was hypothesized to reflect gains in performance due to the intervention.

The statistical analyses tested the main effects of time (T1–T2 gain vs. T2–T3 gain), group (experimental vs wait-list), and grade level (Grade 5 vs Grade 10), as well as their interactions, on written sentence generation skills and text composition. Score distribution was checked first (and separately for each grade level), by inspecting skewness and kurtosis. Between-group differences at the pretest were assessed for each grade level (Grades 5 and 10) by independent t-tests or by Chi square analyses (for nominal variables).

As participants were nested within classes, a multilevel analysis was initially planned to control for random effects at the classroom level (Peugh, 2010). Intraclass correlation (ICC) statistics suggested that very little variance, between 0 and 6%, occurred across classrooms. We thus estimated the design effect (DE) to decide whether we needed to control for clustering at the classroom level. The design effect was computed by the following formula: $DE = 1 + (n_c - 1) * ICC$, where n_c is the average number of subjects per cluster. DEs above 2 indicate violations of the independence assumption that results from nested data and thus suggest the need for multilevel modelling (Peugh, 2010). In our study, all DEs were < 2. Consequently, we performed multiple regression analyses.

Six multiple regressions examined the contribution of grade level (Grades 5 and 10), group (experimental or control), time (T1–T2 or T2–T3 gain), and their interactions (Group × Time, Group × Grade level, Grade level × Time, and the three-way interaction Group × Time × Grade level) on students' gains in sentence generation accuracy, sentence generation fluency, sentence reformulation, text macrostructure, language, and fluency. As prior research highlighted the influence of spelling on text generation skills (Berninger et al., 2011; Limpo et al., 2017; Sumner, Connelly, & Barnett, 2016), spelling skills were also included in all regression models as a

control measure. Moreover, as preliminary analyses revealed pretest differences in text fluency (text generation speed) between the experimental and wait-list groups (see Results), pretest (T1) text fluency was controlled too. Statistical significance was adjusted to $p \leq .008$ (Bonferroni corrections) to control for Type 1 errors. Pairwise comparisons were used to explore the Time × Group and Time × Group × Grade level interactions. Effect sizes were estimated by Cohen's *d* (1988). Follow-up maintenance at 5 weeks from the end of the training (T3) was tested by pairwise samples *t*-tests for the experimental group only, who received the intervention between T1 and T2.

Results

For both grade levels, all measures except spelling scores showed skewness and kurtosis values ≤ 1. Nonnormal distribution was expected in spelling, as the majority of Italian students make almost no spelling errors by the end of primary school. As shown in Table 1, the experimental and wait-list groups did not differ in gender distribution, number of students with learning disabilities, or years in Italy. The only difference observed was age between the fifth graders' experimental and wait-list groups. Differences between the experimental and wait-list groups on pretest scores were examined by *t*-tests. These analyses revealed a significant difference between the experimental and wait-list groups at both grade levels (Grades 5 and 10) in compositional text fluency (i.e., text generation speed). In the fifth-grade sample, the wait-list group was more fluent in writing at the pretest. In the 10th-grade sample, the experimental group showed greater pretest fluency. The 10th graders' experimental and wait-list groups also differed in spelling skills, although the difference only approached statistical significance.

Effectiveness of the intervention: time, group, and grade-level effects

Multiple regressions

The multiple regression models are reported in Table 2 (for dependent measures at the sentence level) and Table 3 (for dependent measures at the text level).

Written sentence generation accuracy Group (B = − 3.15, $p < .005$), time (B = − 2.67, $p < .005$), the two-way interaction Group × Time (B = 7.04, $p < . 001$), and the three-way interaction Grade level × Group × Time (B = − 5.78, $p < . 005$) were significant. The regression model was significant, $F(9,220) = 4.54$, $p < . 001$, accounting for 16% of variance in written sentence generation skills. Figure 2 displays the interaction effects observed for written sentence generation accuracy.

Written sentence generation fluency After Bonferroni corrections were applied, none of the variables accounted for a significant variance in written sentence fluency. The regression model was, however, significant—$F(9,220) = 4.91$, $p < . 001$—accounting for 17% of variance in written sentence generation fluency.

Table 2 Parameters estimates for gains at sentence level with the intervention

Parameter	Gain in sentence generation accuracy			Gain in sentence generation fluency			Gain in sentence reformulation		
	B	SE	t	B	SE	t	B	SE	t
Intercept	4.49	0.97	4.61***	0.71	0.59	1.22	1.50	0.20	7.57***
GradeL	−2.70	1.04	−2.59*	0.31	0.63	0.49	−0.53	0.22	−2.41*
Group	−3.15	0.99	−3.20**	0.41	0.59	0.69	−1.26	0.20	−6.28***
Time	−2.67	0.99	−2.70**	1.26	0.59	2.12*	−1.29	0.20	−6.45***
ZSpelling	−0.04	0.28	−0.15	−0.13	0.17	−0.79	0.03	0.06	0.57
TextFluencyT1	−0.01	0.01	−1.06	−0.01	0.00	−1.71	0.00	0.00	−1.11
Group * Time	7.04	1.37	5.12***	1.30	0.83	−.57	2.25	0.28	8.07***
GradeL * Time	2.06	1.47	1.40	−1.39	0.88	−1.57	0.58	0.30	1.90
GradeL * Group	3.72	1.62	2.30*	0.45	0.97	0.46	0.57	0.33	1.71
GradeL * Group * Time	−5.78	2.16	−2.67**	−1.39	1.30	−1.07	−0.76	0.44	−1.71
R^2	.16			.17			.30		
F	4.54***			4.91***			10.21***		

GradeL = fifth versus 10th grade; Group = experimental versus wait list; Time = time interval T1-T2 versus time interval T2–T3; ZSpelling = spelling Zscores; TextFluencyT1 = mean fluency in text production (words per 15 min) at T1

***$p \le .001$, **$p \le .005$, *$p \le .05$

Table 3 Parameters estimates for gains at text level with the intervention

Parameter	Gain in macrostructure			Gain inlanguage			Gain in text fluency		
	B	SE	t	B	SE	t	B	SE	t
Intercept	2.18	0.25	8.63***	0.86	0.13	6.67***	79.02	11.15	7.09***
GradeL	−0.63	0.27	−2.33*	−0.13	0.14	−0.96	−27.36	11.91	−2.30*
Group	−2.17	0.26	−8.46***	−0.48	0.13	−3.64***	−50.37	11.34	−4.44***
Time	−2.36	0.26	−9.25***	−0.67	0.13	−5.11***	−58.79	11.27	−5.21***
ZSpelling	−0.08	0.07	−1.11	0.01	0.04	0.33	−3.25	3.19	−1.02
TextFluencyT1	0.00	0.00	−2.23*	0.00	0.00	−2.14*	−0.27	0.07	−4.17***
Group * Time	4.60	0.36	12.86***	0.68	0.18	3.75***	77.64	15.77	4.92***
GradeL * Time	1.25	0.38	3.29**	0.07	0.19	0.38	58.38	16.81	3.47**
GradeL * Group	1.20	0.42	2.85**	0.24	0.21	1.12	41.38	18.54	2.23*
GradeL * Group * Time	−1.69	0.56	−3.00**	0.21	0.29	0.73	−49.33	24.77	−1.99*
R^2	.50			.20			.20		
F	24.43***			5.98***			5.90***		

GradeL = Fifth versus 10th grade; Group = Experimental versus Wait list; Time = Time interval T1–T2 versus time interval T2–T3; ZSpelling = Spelling Zscores; TextFluencyT1 = Mean fluency in text production (words per 15 min) at T1

***$p \le .001$, **$p \le .005$, *$p \le .05$

Written sentence reformulation Group (B = −1.26, $p < .001$), time (B = −1.29, $p < .001$), and the two-way interaction Group × Time (B = 2.25, $p < .001$) were significant. The regression model was significant, $F(9,220) = 10.21$, $p < .001$, accounting for 30% of variance in written sentence reformulation skills.
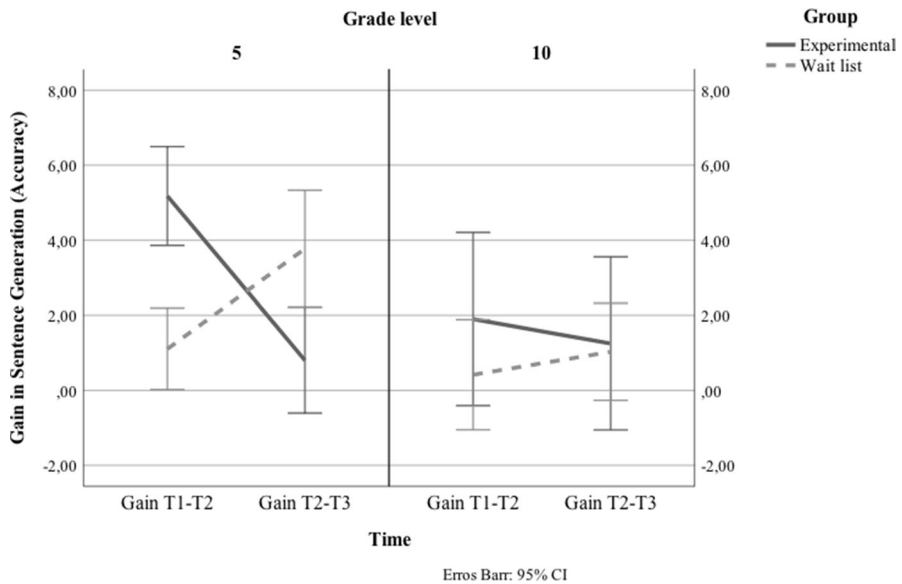
Fig. 2 Effects of the intervention on students' sentence generation accuracy. Note. T1–T2 = training experimental group; T2–T3 = training wait-list group

Text composition quality Two measures of written text quality were considered: macrostructural quality and quality of language.

Macrostructure Group (B = − 2.17, $p <$. 001), time (B = − 2.36, $p <$. 001), the two-way interaction Group × Time (B = 4.60, $p <$. 001), the two-way interactions Grade level × Time (B = 1.25, $p <$. 005) and Grade level × Group (B = 1.20, $p <$. 005), and the three-way interaction Grade level × Group × Time (B = − 1.69, $p <$. 005) were all significant. The regression model was significant, $F(9,220) = 24.42$, $p <$. 001, and accounted for 50% of variance in text macrostructural quality. The interaction effects are displayed in Fig. 3.

Language As for changes in language quality, group (B = − 0.48, $p <$. 001), time (B = − 0.67, $p <$. 001), and the two-way interaction Time × Group (B = 0.68, $p <$. 001), were significant. The regression model was significant, $F(9,220) = 5.98$, $p <$. 001, and accounted for 20% of variance in language quality.

Text composition fluency Group (B = − 50.37, $p <$. 001), time (B = − 58.79, $p <$. 001), the two-way interaction Group × Time (B = 77.64, $p <$. 001), and the two-way interaction Grade level × Time (B = 58.38, $p <$. 005) accounted for significant variance in text writing fluency. The contribution of pretest (T1) text fluency was also significant (B = − 0.27 $p <$. 001). The regression model was significant, $F(9,220) = 5.90$, $p <$. 001, accounting for 20% of variance in text writing fluency.

In synthesis, significant interaction effects between group and time were found for all dependent variables, except for written sentence generation fluency. A three-way interaction, indicating different effects of the intervention depending on grade level
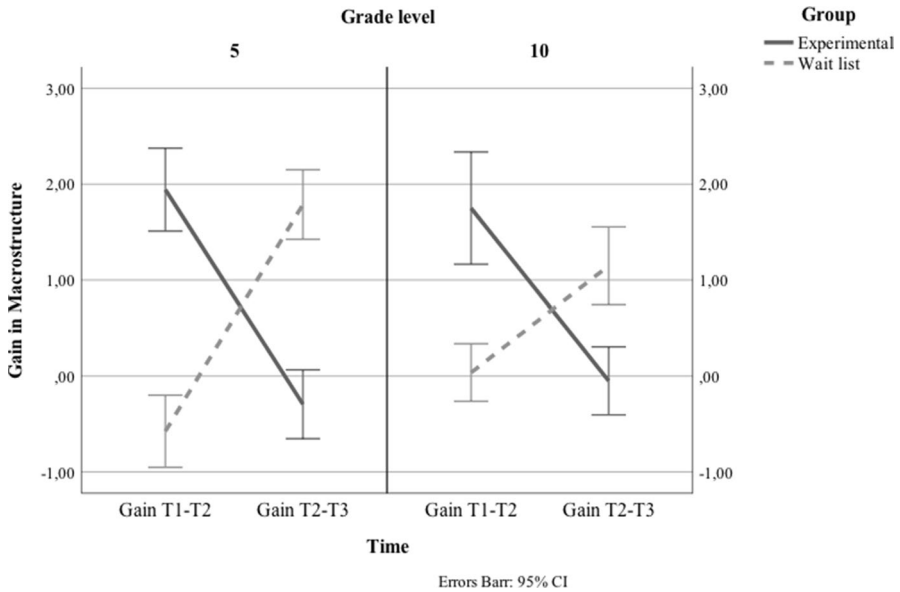
Fig. 3 Effects of the intervention on students' text macrostructure. Note. T1–T2 = training experimental group; T2–T3 = training wait-list group

(Grades 5 and 10), was observed only for written sentence generation accuracy and macrostructural text quality.

Pairwise comparisons

Table 4 displays the mean gains of the two age groups and the experimental and wait-list groups between the first time interval (T1–T2) and the second time interval (T2–T3).

Pairwise comparisons between T1–T2 gain scores and T2–T3 gain scores were run to clarify the two main interactions Time×Group and Time×Group×Grade level. The results for the younger (fifth graders) and older (10th graders) groups are synthetized in Table 5. Table 5 reports the statistical tests and effect size (*d*) of the difference in gain between the two time intervals (T1–T2 gain minus T2–T3 gain). Positive *t*-tests indicate that greater gain occurred between T1 and T2 than between T2 and T3. Negative *t*-tests indicate greater improvement in the second time interval (i.e., T2–T3 gain > T1–T2 gain).

Effects of the intervention on fifth graders' writing skills

As shown in Table 5, the training led to significant improvement for both the experimental and wait-list groups in all the dependent measures, except for text language and text fluency, for which the gain resulted significant only for the

Table 4 Mean gain scores by age group (fifth and 10th graders) and group (experimental and wait list) at the two time intervals (T1–T2 and T2–T3)

	Experimental				Wait list			
	Gain T1–T2		Gain T2–T3		Gain T1–T2		Gain T2–T3	
	M	SD	M	SD	M	SD	M	SD
Grade 5								
Sent_Gen_Acc	5.18	3.84	0.80	4.10	1.11	3.07	3.77	4.40
Sent_Gen_ Fluency	3.16	2.75	0.60	2.89	1.27	2.00	0.02	2.73
Sent_ Reformulation	1.09	0.83	0.12	0.76	0.05	0.67	1.34	0.92
Macrostructure	1.94	1.26	− 0.29	1.03	− 0.58	1.06	1.79	1.02
Language	0.26	0.51	0.24	0.61	0.00	0.35	0.67	0.60
Text fluency (words in 15 min)	23.20	29.47	3.88	44.26	− 12.24	36.27	46.55	40.68
Grade 10								
Sent_Gen_Acc	1.90	4.93	1.25	4.93	0.42	3.71	1.03	3.27
Sent_Gen_Fluency	0.55	2.52	0.77	2.20	0.17	2.02	0.30	1.64
Sent_Reformulation	0.84	0.87	0.05	0.54	0.08	0.85	0.79	0.95
Macrostructure	1.75	1.25	− 0.05	0.76	0.04	0.76	1.15	1.03
Language	0.50	0.69	0.20	0.62	− 0.07	0.38	0.52	0.51
Text fluency (words in 15 min)	20	87.04	− 7.9	63.42	17.07	41.41	17.48	39.96

Gain T1–T2 = gain score between T1 and T2; Gain T2–T3 = gain score between T2 and T3

Table 5 Pairwise comparisons: difference in gains between T1–T2 and T2–T3 by group: t-tests and Cohen's *d*

	Grade 5				Grade 10			
	Experimental		Wait list		Experimental		Wait list	
	t	d	t	d	t	d	t	d
Sent_Gen_Acc	3.91***	0.66	− 2.80**	− 0.49	0.31	0.07	− 0.61	− 0.12
Sent_Gen_Fluency	3.08**	0.52	1.96#	0.34	− 0.22	− 0.05	− 0.23	− 0.04
Sent_Reformulation	4.71***	0.81	− 5.54***	− 0.96	2.81*	0.63	− 1.97#	− 0.41
Macrostructure	6.54***	1.12	− 7.62***	− 1.33	4.64***	1.04	− 3.61**	− 0.69
Language	0.18	0.03	− 4.93***	− 0.86	1.19	0.27	− 5.38***	− 1.03
Text fluency	1.66	0.29	− 5.19***	− 0.90	0.93	0.21	− 0.03	− 0.01

Negative *t*s and *d*s for the wait-list group indicate greater gains between T2 and T3 (after the intervention) than between T1 and T2 (waiting period). Sent_Gen_Acc = mean sentence generation accuracy; Sent_Gen_Fluency = mean sentence generation fluency; Sent_Reformulation = mean sentence reformulation score; Macrostructure = mean text macrostructure score; Language = mean text language score; Text fluency = mean fluency in text production (words per 15 min)

***p ≤ .001,   **p ≤ .005, *p ≤ .05

wait-list group, and for sentence fluency, for which the wait-list group did not show significant improvement. Positive *t* values in Table 5 for the experimental

group and negative t values for the wait-list group indicate that the gain in writing scores at the sentence and text level was greater between T1 and T2 than between T2 and T3 for the experimental group (i.e., T1–T2 gain > T2–T3 gain), and was greater between T2 and T3 than between T1 and T2 for the wait-list group (i.e., T1–T2 gain < T2–T3; see also Figs. 2 and 3). This pattern confirms that improvement occurred as a consequence of the intervention in both groups. The effect sizes ranged from medium ($d = -.49$) to large ($d = -1.33$), with larger effect sizes in both groups on the sentence reformulation tasks ($d = .81$ for the experimental group and $d = -.96$ for the wait-list group) at the sentence level, and on text macrostructure ($d = 1.12$ for the experimental group and $d = -1.33$ for the wait-list group) at the text level. For the wait-list group, the effect size was also large for gains in text language ($d = -.86$) and text fluency ($d = -.90$).

Effects of the intervention on 10th graders' writing skills

For these older writers, the training led to significant improvements in both the experimental and wait-list group only in the macrostructural quality of the texts (see Fig. 3). The effect size was large for the experimental group ($d = 1.04$) and medium for the wait-list group ($d = -.69$). For text language, the training proved to be effective for the wait-list group only (the effect size was large: $d = -1.03$). Although the improvement in sentence reformulation skills was not significant after Bonferroni corrections, the dimension of the effect for the sentence reformulation skills was moderate ($d = .63$ for the experimental group and $d = -.41$ for the wait-list group).

Summing up, the pairwise comparisons confirmed the existence of different training effects for students of different grade levels: The training was less effective in boosting sentence-level language skills for the older writers. By contrast, at the text level, it led to significant improvements for both age groups, especially when macrostructural quality was considered.

Follow-up

For the experimental group, who received the intervention between T1 and T2 (first time interval), we also tested the maintenance of the training effects at a 5-week follow-up. Paired samples t-tests between the students' writing scores (sentence generation accuracy and fluency, sentence reformulation, text macrostructure, text language, and text writing fluency) at T2 and T3 revealed no significant decline in writing scores for any of the dependent variables. Table 4 shows that little change occurred between T2 and T3 for the experimental groups.

Discussion

Oral language is a largely neglected component of instructional writing interventions. In this study, we tested the effectiveness of a nine-session classroom-level writing intervention focused on students' oral sentence construction and reformulation skills. The students involved in this trial had mostly automatized their transcription skills but were still struggling with text generation and oral language abilities, as noted by their teachers and demonstrated by the significant effects of the training.

The results revealed a general effectiveness of oral language training for students' writing, especially for the younger participants (fifth graders), with significant improvements in both the experimental and wait-list groups in all dependent sentence-level writing measures after the intervention—except for sentence fluency. That is, the younger writers transferred the oral sentence generation skills acquired during the training to writing (written sentence generation and reformulation, representing close transfer). Only sentence fluency did not significantly improve for the wait-list group. The benefits of the training generalized for these fifth graders to text composition as well (far transfer), leading to significant improvements in the macrostructural quality of texts produced after the intervention, with large effect sizes ($d = 1.12$ for the experimental and $d = -1.33$ for the wait-list group).

For the older writers (10th graders), the effects of the training were overall not significant at the sentence level. However, these writers significantly benefitted from the intervention at the text level. The oral (sentence generation) intervention led to significant improvements in the macrostructural quality of their texts, with large ($d = 1.04$) and moderate ($d = -.69$) effect sizes for the experimental and wait-list groups, respectively. The results of the training were also significant for text language in the wait-list group only ($d = -1.03$).

Sentences represent the natural processing units of text generation (Berninger et al., 2011; Saddler & Graham, 2005), as ideas are translated into sentences when building a text. Developing the ability to translate ideas in sentences thus represents a foundational writing skill and prerequisite for higher-level coherence-making writing processes. It is thus little surprising that an intervention focused on sentence construction and reformulation led to significant improvements in the macrostructural quality of the texts produced. Similar findings have been obtained in other studies focused on sentence construction (sentence-combining) skills (Limpo & Alves, 2013; Saddler & Graham, 2005). However, these intervention studies did not test the transfer of oral sentence generation skills to writing. Instead, they trained students on written sentence-construction skills directly. Our intervention adds to this important research line by demonstrating that targeting oral language skills at the classroom level can benefit students' writing.

The focus and nature of our training (oral vs written language) can explain the large effect sizes obtained in this study. The magnitude of the training effects (for the younger group in particular: $d = .81$ and $-.96$ for written sentence reformulation in the experimental and wait-list group and $d = 1.12$ and -1.33 for macrostructural quality in the same groups) is larger than that reported by meta-analyses for interventions targeting grammar or sentence structure (ES = 0.50; Graham

& Perin, 2007). Two differences between the present training and those discussed in Graham and Perin's meta-analysis are that: (1) our training was based on oral language activities exclusively. Practicing in oral language, students could focus on idea translation without the additional burdens of transcription. Although relatively fluent in transcription, the participants in this study still made several spelling errors in their texts. This indicated that, under the pressure of text production, transcription skills could still represent a demand for them, as for many struggling writers; (2) all training activities in the present study were group-based and game-like (i.e., playful activities). Other authors have stressed the importance of engaging students in playful and rewarding language/writing activities (Boscolo et al., 2012). Writing is an extremely demanding task, and students must be deeply engaged and motivated to invest their cognitive and linguistic resources and time in this activity.

A finding of this study is that the effectiveness of the intervention was different for the two age groups. Targeting oral sentence generation skills did not lead to significant improvements in the written sentence generation skills of the older writers (10th graders); however, it did impact the development of their text-level writing: text macrostructural and (for the wait-list group) language quality. Table 1 shows that the older students were good at the sentence generation task and, as one would expect, performed better than the fifth graders on the sentence reformulation task as well. Thus, the oral language skills acquired with the training impacted their written production at higher (coherence-making) levels.

Comparing the effectiveness of an experimental intervention for students of different age, grade, or ability levels is important in making practical decisions on when and how to apply that specific intervention in the future (Jones et al., 2013). It is, for example, important to explore how specific instructional strategies or approaches match or can be adapted to the learning needs of various groups of students.

We expected significant effects on writing fluency, but our findings contradicted our initial hypotheses. Writing fluency increased significantly only in the younger wait-list group. As our training had a significant metalinguistic component, it is possible that the students simply spent more time reflecting on their linguistic choices—planning how to translate their ideas in text—at the expense of their writing fluency, which did not improve.

Another unexpected finding of this study was the different efficacy of the training for text language in the two groups (experimental and wait list). In both age groups (fifth and 10th graders), only the wait-list group improved significantly on this measure. Although no significant differences were found between the experimental and wait-list groups at the pretest (except for text fluency and age for the younger writers), three of the younger (fifth grade) students with learning disabilities and three of the older (10th grade) students who had lived in Italy fewer than 6 years were in the wait-list groups (see Table 1). It is possible that greater improvement of the wait-list groups in text language (vocabulary and syntax) was due to the fact that they included these (and perhaps other) students with greater language needs, who benefitted more from the intervention.

Instructional implications

The finding that oral language interventions can lead to improvements in writing has practical implications for the teaching of writing, especially considering that, for many students, poor transcription skills represent a significant barrier to their writing development (Graham & Santangelo, 2014; Hebert, Bohaty et al., 2018, Hebert, Kearns et al., 2018; Sumner et al., 2016). For these students, as well as for those with language-learning needs (Hoff, 2013), as were probably many of the students in this study, practicing text generation processes without the additional burden of transcription may represent an empowering learning experience that could lead to the development of both language and text generation skills.

Oral language activities like those proposed in this study can be particularly beneficial because they are based on immediate peer feedback, collaborative work, and communicative tasks. All these elements may increase students' engagement in the language tasks, stimulating the development of linguistic and metalinguistic skills at the individual level.

The idea that writing instruction can be also realized by oral language activities, especially during the elementary school years, leads to consideration of the role of oral language teaching in teachers' professional training. Viewing oral language as an actual component (not only a prerequisite) of written language instruction necessitates a shift in teaching methods and strategies. Should future studies confirm the effectiveness of oral language interventions for the development of writing skills, teachers' professional training could start integrating oral language components into their programs.

Traditionally, in Europe and the United States, oral language intervention has been considered the field of speech-language pathologists, whose task has been typically to support the oral language underpinnings of reading and writing (Silliman, 2014). The preliminary findings of this study offer a different perspective to these practitioners as well, suggesting that integrated oral language-writing interventions could also be useful in supporting children's language/literacy development.

Limitations and future directions

In assessing intervention effectiveness, we must consider both the strength of an intervention's effects and the internal and external validity of the intervention (McMaster et al., 2018). In this study, the experimenter administered the training directly. Although this approach reduces the risks of low treatment fidelity significantly, it also represents a threat to the external validity of a study, as it is uncertain whether classroom teachers would be able to conduct the intervention themselves. In our study, teachers were involved and assisted the experimenter but were not responsible for conducting the training. Our next step will be to test whether these positive effects can extend to situations in which classroom teachers are in charge of interventions.

Other methodological limitations of this study concern the lack of a precise measure of treatment fidelity for students' participation in the intervention at the individual level, and the high attrition recorded in the study. As for treatment fidelity, we checked that each classroom received instruction for the same length of time and designed the intervention so that all students were forced to take turns leading the teamwork sessions. However, we did not monitor individual students' participation in the training with systematic observations (Ledford et al., 2014). Systematic observations could better ensure that students' engagement in oral language activities was similar among participants and classrooms. As noted in the Participants section, attrition in this study was also high, in particular among our older participants. For students with a low socioeconomic background, like were many of these high school students, school absenteeism can be high (Klein et al., 2020). Unfortunately, it was impossible to make up the missed assessment or training sessions of these students due to scheduling conflicts. This led to a high dropout rate.

A final limitation of this study concerns the long-term efficacy of the intervention. The training was largely effective, especially for younger writers (fifth graders), and its positive effects were maintained 5 weeks after the end of the intervention. However, a long-term follow-up would be needed to conclude that this relatively short (3-week) intervention produced stable gains in students' writing skills. Although long-term follow-ups may have significant costs in terms of dropouts, they represent a unique means to ascertain the long-term efficacy of an intervention, which is what is most important in school.

Conclusions

Writing research has demonstrated that multicomponent interventions combining the strengths of focusing on transcription and text generation (Berninger et al., 2002), or on transcription and self-regulation (Limpo & Alves, 2018), are generally very effective in developing writing skills. The effects of a combined (multi-component) intervention could be further emphasized when students are allowed to focus on (linguistic) text generation processes without the burdens of transcription while developing transcription skills in ad hoc activities. On the other hand, interventions combining self-regulation and (linguistic) text generation processes could be another promising avenue to support the development of high-level writing skills in older writers who have automatized spelling and handwriting. In school, some basic oral language skills, such as phonological skills, are currently overemphasized, while other higher-level language skills, such as vocabulary and syntactic skills, do not seem to receive sufficient attention. Evidence-based instruction of these skills is still lacking.

References

Abbott, R. D., & Berninger, V. W. (1993). Structural equation modeling of relationships among developmental skills and writing skills in primary-grade and intermediate-grade writers. *Journal of Educational Psychology, 85*(3), 478–508. https://doi.org/10.1037/0022-0663.85.3.478.

Abbott, R. D., Berninger, V. W., & Fayol, M. (2010). Longitudinal relationships of levels of language in writing and between writing and reading in Grades 1 to 7. *Journal of Educational Psychology, 102*(2), 281–298. https://doi.org/10.1037/a0019318.

Alamargot, D., Flouret, L., Larocque, D., Caporossi, G., Pontart, V., Paduraru, C., et al. (2015). Successful written subject-verb agreement: An online analysis of the procedure used by students in Grades 3, 5 and 12. *Reading and Writing, 28*(3), 419. https://doi.org/10.1007/s11145-014-9528-x.

Alves, R. A., & Limpo, T. (2015). Progress in written language bursts, pauses, transcription, and written composition across schooling. *Scientific Studies of Reading, 19,* 374–391. https://doi.org/10.1080/10888438.2015.1059838.

Arfe, B., Cona, E., & Merella, A. (2018). Training implicit learning of spelling in Italian children with developmental dyslexia. *Topics in Language Disorders, 38*(4), 299–315. https://doi.org/10.1097/tld.0000000000000163.

Arfe, B., Dockrell, J. E., & De Bernardi, B. (2016). The effect of language specific factors on early written composition: The role of spelling, oral language and text generation skills in a shallow orthography. *Reading and Writing, 29*(3), 501–527. https://doi.org/10.1007/s11145-015-9617-5.

Babayigit, S., & Stainthorp, R. (2010). Component processes of early reading, spelling, and narrative writing skills in Turkish: A longitudinal study. *Reading and Writing, 23*(5), 539–568. https://doi.org/10.1007/S11145-009-9173-Y.

Berninger, V. W. (1999). Coordinating transcription and text generation in working memory during composing: Automatic and constructive processes. *Learning Disability Quarterly, 22*(2), 99–112. https://doi.org/10.2307/1511269.

Berninger, V. W. (2000). Development of language by hand and its connections with language by ear, mouth, and eye. *Topics in Language Disorders, 20*(4), 65–84.

Berninger, V. W., & Abbott, R. D. (2010). Listening comprehension, oral expression, reading comprehension, and written expression: Related yet unique language systems in Grades 1, 3, 5, and 7. *Journal of Educational Psychology, 102*(3), 635. https://doi.org/10.1037/a0019319.

Berninger, V. W., Nagy, W., & Beers, S. (2011). Child writers' construction and reconstruction of single sentences and construction of multi-sentence texts: Contributions of syntax and transcription to translation. *Reading and Writing, 24*(2), 151–182. https://doi.org/10.1007/s11145-010-9262-y.

Berninger, V. W., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Curtin, G., et al. (2002). Teaching spelling and composition alone and together: Implications for the simple view of writing. *Journal of Educational Psychology, 94*(2), 291. https://doi.org/10.1037//0022-0663.94.2.291.

Berninger, V., Yates, C., Cartwright, A., Rutberg, J., Remy, E., & Abbott, R. (1992). Lower-level developmental skills in beginning writing. *Reading and Writing, 4*(3), 257–280. https://doi.org/10.1007/bf01027151.

Boscolo, P., Gelati, C., & Galvan, N. (2012). Teaching elementary school students to play with meanings and genre. *Reading and Writing Quarterly, 28*(1), 29–50. https://doi.org/10.1080/10573569.2012.632730.

Brimo, D., & Hall-Mills, S. (2019). Adolescents' production of complex syntax in spoken and written expository and persuasive genres. *Clinical Linguistics and Phonetics, 33*(3), 237–255. https://doi.org/10.1080/02699206.2018.1504987.

Campbell, M. J., Hemming, K., & Taljaard, M. (2019). The stepped wedge cluster randomised trial: What it is and when it should be used. *Medical Journal of Australia, 210*(6), 253. https://doi.org/10.5694/mja2.50018.

Chanquoy, L., & Negro, I. (1996). Subject-verb agreement errors in written productions: A study of French children and adults. *Journal of Psycholinguistic Research, 25*(5), 553–570. https://doi.org/10.1007/bf01758183.

Cohen, J. (1988). *Statistical power analysis for the behavioral science* (2nd ed.). New Jersey: Lawrence Erlbaum.

Danzak, R. L., & Arfe, B. (2016). Globally minded text production bilingual, expository writing of Italian adolescents learning English. *Topics in Language Disorders, 36*(1), 35–51. https://doi.org/10.1097/tld.0000000000000077.

Datchuk, S. M., & Kubina, R. M. (2013). A review of teaching sentence-level writing skills to students with writing difficulties and learning disabilities. *Remedial and Special Education, 34*(3), 180–192. https://doi.org/10.1177/0741932512448254.

Dockrell, J. E., & Connelly, V. (2013). The role of oral language in underpinning the text generation difficulties in children with specific language impairment. *Journal of Research in Reading.* https://doi.org/10.1111/j.1467-9817.2012.01550.x.

Dockrell, J. E., Connelly, V., & Arfe, B. (2019). Struggling writers in elementary school: Capturing drivers of performance. *Learning and Instruction, 60,* 75–84. https://doi.org/10.1016/j.learninstruc.2018.11.009.

Dockrell, J. E., Lindsay, G., & Connelly, V. (2009). The impact of specific language impairment on adolescents' written Text. *Exceptional Children, 75*(4), 427–446. https://doi.org/10.1177/001440290907500403.

Dockrell, J. E., Lindsay, G., Connelly, V., & Mackie, C. (2007). Constraints in the production of written text in children with specific language impairments. *Exceptional Children, 73*(2), 147–164.

Gamez, P. B., Lesaux, N. K., & Rizzo, A. A. (2016). Narrative production skills of language minority learners and their English-only classmates in early adolescence. *Applied Psycholinguistics, 37*(4), 933–961. https://doi.org/10.1017/s0142716415000314.

Goldfeld, S., Snow, P., Eadie, P., Munro, J., Gold, L., Le, H. N. D., et al. (2017). Classroom promotion of oral language (CPOL): Protocol for a cluster randomised controlled trial of a school-based intervention to improve children's literacy outcomes at Grade 3, oral language and mental health. *British Medical Journal Open, 7*(11), e016574. https://doi.org/10.1136/bmjopen-2017-016574.

Graham, S., Berninger, V. W., Abbott, R. D., Abbott, S. P., & Whitaker, D. (1997). Role of mechanics in composing of elementary school students. A new methodological approach. *Journal of Educational Psychology, 89*(1), 170. https://doi.org/10.1037/0022-0663.89.1.170.

Graham, S., Hebert, M., Fishman, E., Ray, A. B., & Rouse, A. G. (2020). Do children classified with specific language impairment have a learning disability in writing? A meta-analysis. *Journal of Learning Disabilities, 53*(4), 292–310. https://doi.org/10.1177/0022219420917338.

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. https://doi.org/10.1037/a0029185.

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445–476. https://doi.org/10.1037/0022-0663.99.3.445.

Graham, S., & Santangelo, T. (2014). Does spelling instruction make students better spellers, readers, and writers? A meta-analytic review. *Reading and Writing, 27*(9), 1703–1743. https://doi.org/10.1007/s11145-014-9517-0.

Gugliotta, M., Bisiacchi, P. S., Cendron, M., Tressoldi, P. E., & Vio, C. (2009). *BVN 12-18: Bateria di valutazione neuropsicologica per l'adolescenza [Battery for neuropsychological assessment in adolescence].* Trento: Erikson.

Hebert, M., Bohaty, J. J., Nelson, J. R., & Roehling, J. V. (2018). Writing informational text using provided information and text structures: An intervention for upper elementary struggling writers. *Reading and Writing, 31*(9), 2165–2190. https://doi.org/10.1007/s11145-018-9841-x.

Hebert, M., Kearns, D. M., Hayes, J. B., Bazis, P., & Cooper, S. (2018). Why children with dyslexia struggle with writing and how to help them. *Language Speech and Hearing Services in Schools, 49*(4), 843–863. https://doi.org/10.1044/2018_lshss-dyslc-18-0024.

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology, 49*(1), 4–14. https://doi.org/10.1037/a0027238.

Hulme, C., & Snowling, M. J. (2014). The interface between spoken and written language: Developmental disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1634), 20120395. https://doi.org/10.1098/rstb.2012.0395.

Jennifer, R. L, Ayres, K. M., Lane, J. D., Lam, M. F. (2014) Identifying Issues and Concerns With the Use of Interval-Based Systems in Single Case Research Using a Pilot Simulation Study. *The Journal of Special Education 49*(2):104–117. https://doi.org/10.1177/0022466915568975

Jones, S., Myhill, D., & Bailey, T. (2013). Grammar for writing? An investigation of the effects of contextualised grammar teaching on students' writing. *Reading and Writing, 26*(8), 1241–1263. https://doi.org/10.1007/s11145-012-9416-1.

Juel, C. (1988). Learning to read and write—A longitudinal study of 54 children from 1st through 4th grades. *Journal of Educational Psychology, 80*(4), 437–447. https://doi.org/10.1037/0022-0663.80.4.437.

Kent, S., Wanzek, J., Petscher, Y., Al Otaiba, S., & Kim, Y. S. (2014). Writing fluency and quality in kindergarten and first grade: The role of attention, reading, transcription, and oral language. *Reading and Writing, 27*(7), 1163–1188. https://doi.org/10.1007/s11145-013-9480-1.

Kim, Y. S. G. (2016). Direct and mediated effects of language and cognitive skills on comprehension of oral narrative texts (listening comprehension) for children. *Journal of Experimental Child Psychology, 141,* 101–120. https://doi.org/10.1016/j.jecp.2015.08.003.

Kim, Y. S. G. (2019). Structural relations of language and cognitive skills, and topic knowledge to written composition: A test of the direct and indirect effects model of writing. *British Journal of Educational Psychology*. https://doi.org/10.1111/bjep.12330.

Kim, Y. S., Al Otaiba, S., Wanzek, J., & Gatlin, B. (2015). Toward an understanding of dimensions, predictors, and the gender gap in written composition. *Journal of Educational Psychology, 107*(1), 79–95. https://doi.org/10.1037/a0037210.

Kim, Y. S. G., Gatlin, B., Al Otaiba, S., & Wanzek, J. (2018). Theorization and an empirical investigation of the component-based and developmental text writing fluency construct. *Journal of Learning Disabilities, 51*(4), 320–335. https://doi.org/10.1177/0022219417712016.

Kim, Y. S. G., & Schatschneider, C. (2017). Expanding the developmental models of writing: A direct and indirect effects model of developmental writing (DIEW). *Journal of Educational Psychology, 109*(1), 35–50. https://doi.org/10.1037/edu0000129.

Klein, M., Sosu, E., & Dare, S. (2020). Mapping inequalities in school attendance: the relationship between different dimensions of socioeconomic status and forms of school absence. *Children and Youth Services Review, 118,* 105432. https://doi.org/10.31235/osf.io/9jrqx.

Limpo, T., & Alves, R. A. (2013). Teaching planning or sentence-combining strategies: Effective SRSD interventions at different levels of written composition. *Contemporary Educational Psychology, 38*(4), 328–341. https://doi.org/10.1016/j.cedpsych.2013.07.004.

Limpo, T., & Alves, R. A. (2018). Tailoring multicomponent writing interventions: Effects of coupling self-regulation and transcription training. *Journal of Learning Disabilities, 51*(4), 381–398. https://doi.org/10.1177/0022219417708170.

Limpo, T., Alves, R. A., & Connelly, V. (2017). Examining the transcription-writing link: Effects of handwriting fluency and spelling accuracy on writing performance via planning and translating in middle grades. *Learning and Individual Differences, 53,* 26–36. https://doi.org/10.1016/j.lindif.2016.11.004.

McCutchen, D., Covill, A., Hoyne, S. H., & Mildes, K. (1994). Individual differences in writing—Implications of translating fluency. *Journal of Educational Psychology, 86*(2), 256. https://doi.org/10.1037//0022-0663.86.2.256.

McMaster, K. L., Kunkel, A., Shin, J., Jung, P. G., & Lembke, E. (2018). Early writing intervention: A best evidence synthesis. *Journal of Learning Disabilities, 51*(4), 363–380. https://doi.org/10.1177/0022219417708169.

Mehta, P. D., Foorman, B. R., Branum-Martin, L., & Taylor, W. P. (2005). Literacy as a unidimensional multilevel construct: Validation, sources of influence, and implications in a longitudinal study in grades 1 to 4. *Scientific Studies of Reading, 9*(2), 85–116.

Myhill, D. A., Jones, S. M., Lines, H., & Watson, A. (2012). Re-thinking grammar: the impact of embedded grammar teaching on students' writing and students' metalinguistic understanding. *Research Papers in Education, 27*(2), 139–166. https://doi.org/10.1080/02671522.2011.637640.

Olive, T., & Kellogg, R. T. (2002). Concurrent activation of high- and low-level production processes in written composition. *Memory and Cognition, 30*(4), 594–600. https://doi.org/10.3758/bf03194960.

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48,* 85–112. https://doi.org/10.1016/j.jsp.2009.09.002.

Pinto, G., Tarchi, C., & Bigozzi, L. (2015). The relationship between oral and written narratives: A three-year longitudinal study of narrative cohesion, coherence, and structure. *British Journal of Educational Psychology, 85*(4), 551–569. https://doi.org/10.1111/bjep.12091.

Ravid, D., & Berman, R. A. (2006). Information density in the development of spoken and written narratives in English and Hebrew. *Discourse Processes, 41*(2), 117–149. https://doi.org/10.1207/s15326950dp4102_2.

Ravid, D., & Tolchinsky, L. (2002). Developing linguistic literacy: A comprehensive model. *Journal of Child Language, 29*(2), 417–447. https://doi.org/10.1017/s0305000902005111.

Rust, J. (1996). *The manual of the Wechsler objective language dimensions (WOLD)*. New York: The Psychological Corporation.

Saddler, B., Behforooz, B., & Asaro, K. (2008). The effects of sentence-combining instruction on the writing of fourth-grade students with writing difficulties. *Journal of Special Education, 42*(2), 79–90. https://doi.org/10.1177/0022466907310371.

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. https://doi.org/10.1037/0022-0663.97.1.43.

Sartori, G., Job, R., & Tressoldi, P. (2007). *DDE-2. Batteria per la valutazione della dislessia e della disortografia evolutiva [Battery for the assessment of developmental dyslexia and dysorthographia]*. Firenze: Giunti OS.

Schneider, J. M., Abel, A. D., Ogiela, D. A., Middleton, A. E., & Maguire, M. J. (2016). Developmental differences in beta and theta power during sentence processing. *Developmental Cognitive Neuroscience, 19,* 19–30. https://doi.org/10.1016/j.dcn.2016.01.001.

Silliman, E. R. (2014). Integrating oral and written language into a new practice model. In B. Arfé, J. Dockrell, & V. W. Berninger (Eds.), *Writing development in children with hearing loss, dyslexia, or oral language problems*. Oxford: Oxford University Press.

Spencer, T. D., & Petersen, D. B. (2018). Bridging oral and written language: An oral narrative language intervention study with writing outcomes. *Language Speech and Hearing Services in Schools, 49*(3), 569–581. https://doi.org/10.1044/2018_lshss-17-0030.

Sumner, E., Connelly, V., & Barnett, A. L. (2016). The influence of spelling ability on vocabulary choices when writing for children with dyslexia. *Journal of Learning Disabilities, 49*(3), 293–304. https://doi.org/10.1177/0022219414552018.

Tesnière, L. (2015). *Elements of structural syntax* (T. Osborne and S. Kahane, Trans.). John Benjamins. (Original work published 1965).

Wijekumar, K., Graham, S., Harris, K. R., Lei, P. W., Barkel, A., Aitken, A., et al. (2019). The roles of writing knowledge, motivation, strategic behaviors, and skills in predicting elementary students' persuasive writing from source material. *Reading and Writing, 32*(6), 1431–1457. https://doi.org/10.1007/s11145-018-9836-7.