



The Effect of Sentence Combining Instruction with Second- to Fourth-Grade Children: a Replication Study in Turkey

Omer Faruk Tavsanlı¹ · Steve Graham² · Abdullah Kaldirim³ · Alyson Collins⁴

Accepted: 22 August 2023 / Published online: 13 September 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023, corrected publication 2023

Abstract

Translating ideas into acceptable sentence is an essential writing production process. Limited sentence construction skills can hinder young writers from expressing ideas as intended or creating sentences that are comprehensible to their audiences. This may also limit other writing production processes, as young writers must devote considerable attention to this skill until it becomes more facile. This investigation replicated an earlier sentence combining study conducted in the USA by Saddler and Graham (Journal of Educational Psychology, 97:43–54, 2005). In the current study, 88 Grade 2 to 4 Turkish students who received sentence combining instruction that included peer-assisted learning were compared to 83 students in the same grades and school who continued to receive their regular classroom writing instruction. Students receiving sentence combining instruction had statistically higher scores on measures of sentence fluency, writing quality, and length of essays than students in the business-as-usual comparison. The study provided evidence that the peer-assisted learning model of sentence combining instruction tested here and in Saddler and Graham (Journal of Educational Psychology, 97:43–54, 2005) was effective. These findings also provided support for the importance of sentence construction skills, as teaching such skills resulted in more general improvements in writing, including an improvement of overall quality of text. Implications for practice, theory, and research are discussed.

Keywords Writing · Sentence combining · Sentence construction · Sentence fluency · Writing processes

✉ Steve Graham
steve.graham@asu.edu

¹ Istanbul Aydin University, Istanbul, Turkey

² Arizona State University, Tempe, AZ, USA

³ Kutahya Dumlupinar University, Kutahya, Turkey

⁴ Texas State University, San Marcos, TX, USA

An essential component of writing in Hayes' (1996) influential model of writing was text production. He proposed a provisional model of text production where writers draw upon cues from their plans for writing and the text produced so far to retrieve or acquire pertinent semantic content. This content is then converted in working memory into written sentences.

Text production processes were also central to the Writer(s)-Within-Community model (WWC; Graham, 2018), which acted as the theoretical foundation for the current study. According to this model, writing is interactively shaped and bound by the communities in which it takes place and the cognitive capabilities and resources of writers within these communities. The cognitive processes of text production in the WWC model include conceptualization (mentally representing the writing task through structures like goals, written plans, diagrams, pictures, and text produced so far), ideation (acquiring from long-term memory or external sources language, abstract thoughts, images, or sounds for possible writing content), translation (turning selected writing content into acceptable sentences that meet the writer's intention), transcription (converting sentence parts or sentences the writer creates into written or digital text), and reconceptualization (rethinking or revising writing conceptualization, ideation, translation, and transcription).

The production processes specified in the WWC model (Graham, 2018) are initiated and orchestrated through executive control processes a writer commands (e.g., formulating intentions) and draws upon resources in long-term memory (e.g., schemas for writing a story, paragraph, or sentence), are impacted by beliefs writers hold about the writing task (e.g., view it as valuable) and the community in which writing is produced (e.g., share the goals of the community), and are executed through deliberate mental labor in working memory. Translation, which is the focus of this investigation and the production process most closely aligned with Hayes' (1996) description of text production, is conscious, controllable, involves serial processing, and is limited by attentional and working memory resources (Graham, 2021). In both models, the construction of sentences can be more or less effortful depending on the language, images, or abstract thoughts writers are trying to translate into writing and the richness of their knowledge about sentence structure, grammar, usage, and vocabulary.

For young developing writers, mastering the production process of translation is an important step in becoming a capable writer (Limpo & Alves, 2013). Until they gain control over the sentence construction process, this skill can hinder their ability to translate ideas into their intended meanings (Saddler, 2012). It may also inhibit or even interfere with other writing production processes (Graham, 2018). If a child has to devote considerable effort to crafting a sentence, this can deplete the cognitive resources that are available for other production processes such as ideation. Moreover, interference can occur when a child has to devote considerable attention or time to determining how to craft a particular sentence, as plans or ideas held in working memory may be lost or forgotten. (Graham, 2006)

One approach for making the translation skills of developing writers more facile is sentence combining (the focus of the present investigation). This method involves teaching students how to construct more complex and sophisticated sentences by combining two or more basic (i.e., kernel) sentences into a single syntactically correct sentence (Saddler, 2012). Sentence combining is designed to help students become

more skilled at translating ideas into sentences that are understandable to readers. The assumed benefits of sentence combining instruction rest on three assumptions. One, such instruction provides students with greater mastery with schemas or sentence structures they can use in a facile but deliberate and effective manner, which in turn can reduce the cognitive load young writers' experience during translation (Saddler & Graham, 2005). Two, once the production process of sentence construction becomes more habitual as a result of sentence combining instruction, there should be improvements in other aspects of students' writing including the quality and quantity of what they write (Graham, 2021). Three, sentence combining involves integrating already crafted kernel sentences into more complex sentences, removing the necessity of generating ideas or selecting words to represent those ideas. This allows students to focus more specifically on mastering the underlying schemas for creating more sophisticated sentences. The sentence combining instruction applied in this investigation, however, provided students with a bridge for transferring their newly learned sentence skills to writing, as they also practiced using these skills when drafting and revising text.

According to the WWC model (Graham, 2018), the teaching of writing, including the teaching of sentence construction skills or other production processes, is influenced by the communities within which it takes place. To illustrate, the writing instruction that occurs in a specific classroom and its effectiveness depends on a variety of factors, including the purposes writing is meant to achieve, types of writing valued, established writing practices, tools available for writing, social and physical arrangement of the environment, and collective history of the classroom. These attributes of a writing community are in turn shaped by cultural, social, institutional, political, and historical determinants operating outside the classroom. Consequently, the effectiveness of specific instructional practices for teaching writing are likely to differ in contexts that are dissimilar (e.g., Hsiang & Graham, 2016; Hsiang et al., 2018). In the current study, we tested the effectiveness of providing sentence combining instruction to Turkish students in Grades 2 to 4.

Meta-analyses and systematic reviews summarizing the outcomes from writing instructional studies, including sentence combining studies, involved investigations that were mainly conducted in the USA and Europe (e.g., Andrews et al., 2006; Graham et al., 2012; Graham & Perin, 2007; Koster et al., 2015). If we are to gain a better understanding of whether these methods are applicable in different countries, especially ones that have different cultural, social, historical, and political trajectories, researchers need to conduct replication studies in countries like Turkey. This makes it possible to determine if writing treatments like sentence combining are effective broadly or if the observed effects are more localized.

Turkey provides an interesting context for testing if a writing treatment like sentence combining is effective in countries outside the USA and Europe. Like these countries, Turkey has a secular constitution, but it is situated in an Islamic culture. In comparison to the USA and many European countries, schools in Turkey have little autonomy, as education is highly centralized and educational policy is steered by the Ministry of National Education and the Council of Higher Education (<http://www.oecd.org/education/highlightsturkey.htm>). The scores of Turkish students on OECD reading, math, and science assessments are below average. Graduation rates in Turkey are lower and the proportion of underperforming students is higher in comparison

to Turkey's OECD partners. As a result, education in Turkey lags behind most other OECD countries (Kamal, 2017), and this may also be the case for teaching writing. For close to 20 years, teachers have been encouraged to apply a process-based writing approach; however, some teachers in Turkey apply this approach poorly. Furthermore, up to 70% of teachers continue to apply the traditional product-based approach, emphasizing skills like grammar instruction (e.g., Aşıkcan & Pilten, 2016; Tavşanlı, 2017; Türkben, 2021). These and other factors may make sentence combining instruction more or less effective in Turkey than it is in the USA and Europe.

Effectiveness of Sentence Combining Instruction

Multiple reviews have examined the effectiveness of sentence combining instruction, but not all of them have included studies conducted with elementary grade children. This was the case for Hillocks' (1986) meta-analysis. He identified five sentence combining studies conducted in the USA with students in Grades 7 to college. All of these investigations tested sentence combining instruction with an experimental or quasi-experimental design. These five studies resulted in a statistically significant effect size of 0.35 for writing quality.

Andrews and colleagues (2006) took a different approach to examining the effectiveness of sentence combining. They identified 18 studies that assessed the effectiveness of sentence combining using experimental or quasi-experimental designs. All studies were conducted in the USA (17) or Canada (1). They evaluated the quality of each of these studies and identified four investigations rated as high- or medium-quality investigations. These studies involved either fourth- or seventh-grade students. Based on these four studies, they concluded that sentence combining improved grammar accuracy and writing quality.

The Graham and Perin (2007) located five sentence combining studies conducted with students in Grades 4 to 9. These studies were all conducted in the USA (4) or Europe (1), and they each applied either an experimental or quasi-experimental design. As with the previous reviews, sentence combining instruction in these investigations had a positive impact on the writing quality, resulting in an effect size of 0.50.

In a review of experimental and quasi-experimental writing intervention studies conducted only with students in Grades 4 and 5, Koster et al. (2015) located two sentence combining studies. One was conducted in the USA with fourth-grade students, whereas the other was a European study with students in the same grade. In the European study (Gein, 1991), sentence combining instruction had a small impact on writing quality (effect size = 0.11), whereas the US study by the Saddler and Graham (2005) sentence combining had a large impact (effect size = 1.66).

Several studies not included in the reviews summarized above have examined the effectiveness of sentence combining instruction with elementary school children. In a study conducted in Europe, Limpo and Alves (2013) tested whether providing sentence combining instruction to Grade 5 and 6 students was more effective than providing standard writing instruction. Sentence combining instruction enhanced students' sentence skills (effect size = 1.06), quality of writing (effect size = 0.65), and length of text (effect size = 0.96).

In another study conducted in Europe, Walters et al. (2021) examined if sentence combining instruction improved the writing of students in Grades 2 to 5. All students in this investigation were children who struggled with learning to write. Students were randomly assigned to the treatment and wait list control condition. Sentence combining instruction was provided to small groups as a Tier-2 intervention. Sentence combining improved both sentence skills and writing quality (effect sizes ranged from 0.48 to 0.84).

Collectively, the reviews and studies summarized above demonstrated that sentence combining when conducted with students in the USA and Europe can not only improve sentence skills, but other aspects of composing such as writing quality. These findings are consistent with the theoretical proposition that once sentence construction becomes more habitual as a result of sentence combining instruction, students can allocate cognitive resources to other aspects of writing (Graham, 2021). It must be noted that only one study has applied an experimental design to test the effects of sentence combining instruction with students below Grade 4 (Walters et al., 2021). This study involved students experiencing challenges learning to write, and the students received the instruction in small groups. Further, we were unable to locate any studies testing the effectiveness of sentence combining with elementary grade students in Turkey. The study reported here addressed each of these issues as it used a randomized cluster design to test the effectiveness of sentence combining with typically developing Turkish students in Grades 2 to 4.

Research Questions and the Present Study Replication

The current study was designed to answer the following two questions:

1. Does sentence combining instruction improve students' sentence skills? (RQ1)
2. Does sentence combining instruction increase the quality and length of students' opinion essays? (RQ2)

The sentence combining instruction provided to Turkish students in this study was based on procedures applied with Grade 4 students in the USA by Saddler and Graham (2005). This experimental study was identified by Andrews and his colleagues (2006) as one of two investigations out of the 18 they examined that could be classified as a high-quality investigation. It is also the only sentence combining study to date that makes peer-assisted learning an integral part of the learning process. In this previous investigation, instructors first modeled how to combine targeted compound or complex sentences from simpler kernel sentences, followed by fourth-grade students practicing combining similar kernel sentences and then each student working with a peer to continue this practice, with each peer taking turns as learner and coach. The peers also worked together to apply the sentence combining skills they were learning in the context of writing and revising text. Students in this investigation were more and less capable writers. Less capable writers scored one standard deviation or more below the mean of the normative sample on a standardized test of sentence combining skills (*Test of Written Language-3*; Hammill & Larsen, 1996),

whereas the more capable writers scored at the mean or higher on the same test. All instruction took place outside of the context of the regular classroom.

To assess the effects of sentence combining instruction, Saddler and Graham (2005) tested students' sentence construction skills with a norm-referenced standardized test and examined the quality and length of their stories before and after the end of instruction. The performance of students who received sentence combining instruction was compared to students who were taught grammar. Students in the sentence combining condition made greater sentence construction gains than control students (effect size=0.81) and produced stories of higher quality when revising them at posttest (effect size=0.64). No statistically detectable differences were noted for length of stories.

The present study replicated Saddler and Graham (2005) but investigated whether the sentence combining procedures used in the previous study would be effective with typical students in general education classrooms (the prior study involved small group instruction with more and less capable writers) and Grade 2 to 4 students (the prior study involved only Grade 4 students). Other differences included the following: Sentence skills were assessed within the context of students' writing (the prior study assessed sentence construction using a contrived task), transfer to other aspects of writing were assessed through opinion writing (the prior study focused on story writing), and a business-as-usual (BAU) control condition was applied (the prior study created a grammar instructional control condition). We decided to apply a BAU control condition for the following reason. Writing lessons in Grades 2 to 4 in Turkey place a heavy, but not sole emphasis on teaching grammar (Babayiğit & Stainthorp, 2010; Tavşanlı & Kara, 2022). Thus, the BAU control condition provided a comparison where students were taught grammar, but this was not the sole focus of writing instruction. In our opinion, this provided a more realistic assessments of the possible effects of sentence-combining instruction in the context of typical writing practices in Turkey.

The ultimate test of whether sentence combining instruction enhances sentence construction skills is to determine if there is an improvement in the use of these skills when students write (this was not done in the prior study). Unlike the previous investigation by Saddler and Graham (2005), we asked students to write opinion essays instead of stories because Turkish students profess a preference for writing such text (Seban, 2016; Tavşanlı, 2018). This increased the possibility that students would do their best writing when we assessed transfer effects of sentence combining instruction. In addition, writing opinion essays to convey thoughts and feelings is a curricular objective in the Turkish Writing Curriculum for children in Grades 2 to 4.

One additional difference between the current study and Saddler and Graham (2005) was that treatment students in this investigation were provided instruction on the structure of basic and more complex sentences before they were introduced to sentence combining instruction. This included the use of activities designed to ensure that students could distinguish between well-designed and poorly designed sentences. This instruction was included in this study to ensure that all treatment students, even the youngest ones, understood how sentences were constructed and what constituted a good sentence, providing them with a solid foundation for benefiting from the sentence combining instruction provided.

We predicted that sentence combining instruction would improve students' sentence skills when writing (RQ1) because it provided them with new skills for translating their ideas for text into acceptable sentences. We further predicted that sentence combining instruction would enhance the quality of students' opinion essays (RQ2). Such instruction should improve sentence skills, which in turn should reduce cognitive load when writing (Graham, 2018), freeing up resources that can be applied to other important aspects of composing. We made no prediction concerning the impact of sentence combining instruction on length of students' text. Previous studies have yielded inconsistent findings with regard to improvements in text length (Limpo & Alves, 2013; Saddler & Graham, 2005).

Lastly, we anticipated that students in upper grades would demonstrate better sentence skills when writing, produce higher quality papers, and write longer text than students in lower grades. No predictions were made concerning the differential effectiveness of sentence combining instruction at the three different grade levels. While it is possible that the instruction provided might be more effective with older students than younger ones because it concentrated on compound and complex sentence, such effects may have been mitigated by making sure all treatment students were familiar with simple as well as more sophisticated sentence structures before receiving sentence combining instruction.

Methods

Setting

The study took place in a school located in Istanbul, Turkey. This is the largest city in Turkey, and it serves as the social, cultural, historical, and economic hub of the country. The school was located on the European side of the city, where the social economic status of most families can be described as middle to upper class. The principal and teachers in the participating school emphasized the importance of educational reform and supported projects designed to enhance students' academic, mental, social, and emotional engagement. Families of children at the school commonly expressed enthusiasm for such innovations.

Six second- to fourth-grade teachers in the cooperating school agreed to take part in the present study. This included two teachers in each grade, who were randomly assigned by grade to either the sentence combining treatment or the BAU control condition. All six teachers were certified in primary education and had completed a Bachelor's degree. They also had considerable teaching experience (19 to 25 years of teaching experience each), and each teacher had attended at least five in-service preparations devoted to the teaching of writing. Four of the teachers were women, whereas the two fourth grade teachers were men.

Teachers in the treatment condition agreed not to share sentence combining materials or methods with control teachers. Both groups of teachers were told that these materials and methods would be shared with control teachers once the study was completed. Teachers in Grades 2 to 4 in the participating schools typically devoted 120 min a week to writing and writing instruction. Their writing program was based

on a textbook series that addressed the curriculum objectives for writing set forth by Ministry of National Education in Turkey, which included teaching grammar.

Participants

Teachers sent consent letters to parents in their classes which described the study, possible risks, and advantages for children participating in the study and established that students could withdraw from the study at any point during the experiment. Parents of all 171 students in the six classes granted written permission for their child to participate in the study.

Thirty-four percent of the 171 participating students were in Grade 2 ($n = 58$), 28% in Grade 3 ($n = 48$), and 38% in Grade 4 ($n = 65$). Across Grades 2 to 4, 88 students received the sentence combining treatment, whereas 83 students were in the BAU control classrooms. Slightly less than one half of all students were girls ($n = 83$), with 42 girls in the treatment condition and 41 girls in the BAU classrooms. None of the students in the six classes had been identified as having a disability or a special need. This was determined through interviews with deputy principal and the six teachers in the cooperating school.

Sentence Combining Instruction

General Procedures

Students in the treatment group received 30 h of sentence specific instruction over a 10-week period (three lessons per week, with each lasting 1 h). During the first 2 weeks, students participated in lessons that taught them about basic sentence structures (e.g., simple, compound, and complex) and provided them with practice distinguishing between well-constructed and poorly constructed sentences. The remaining 8 weeks of instruction involved sentence combining instruction, where students were taught how to construct more complex sentence by combining smaller kernel sentences together. This approach to sentence combining was based on the procedures applied by Saddler and Graham (2005) and included peer-assisted instruction.

Students' regular classroom teacher delivered the 2 weeks of sentence and 8 weeks of sentence combining instruction. Before the start of the study, these teachers received 16 h of instruction from the first author of this study on how to implement the sentence combining treatment in eight, 2-h sessions. Teachers were provided with all training materials, which included detailed instructions for delivering sentence and sentence combining instruction activities as well as instructional materials to be used in class. The first author and the teachers discussed, modified as needed, and practiced the activities to be applied by teachers in their classroom.

Treatment teachers received 10 h of ongoing coaching as the study was underway, meeting with the first author for 1 h each week. These 1-h sessions served as both review and practice sessions where teachers and the first author discussed the procedures to be applied during the week, practiced specific instructional activities as needed, and considered issues or problems that might arise during instruction.

Sentence Instruction

The first two weeks of instruction focused on making sure students were familiar with the structure of different kinds of sentences: declarative (makes a statement), interrogative (asks a question), and exclamative (expresses strong emotion or surprise). This instruction was designed by the researchers and was part of the overall sentence combining instructional package. The greatest emphasis during instruction was placed on declarative sentences. Sentence instruction during these first 2 weeks also concentrated on helping students become more adept at distinguishing between well-designed and poorly designed sentences, emphasizing the importance of using well-designed sentences when writing. This goal of this instruction was to ensure that students in the treatment group had a basic understanding of different kinds of sentences and why it was important to use well-designed sentences when writing, such as the ones they would be taught during sentence combining.

The six sentence instruction lessons followed a basic structure. First, students were introduced to examples that represented the three different types of sentences (declarative, interrogative, and exclamative). These included simple, compound, and complex sentences over the course of the six lessons. Students were asked to examine the target sentences closely to determine how they were similar and different. With assistance from their teacher, they identified the structural differences between the target sentences. Students were further asked to identify which of the target sentences presented best conveyed the author's meaning and why this was the case. Next, the teacher removed connecting words (e.g., and) from sentences compound and complex sentences, and students discussed how this influenced sentence quality and meaning. Additionally, students played a game where they rated the quality of different kinds of sentences (well-formed and less-well formed). They compared the scores they gave to sentences used in the game and discussed which of these sentences they wanted to use when writing their own text. The content of sentences during sentence instruction were drawn from learning at school, reading, sports, museums, and shopping.

Sentence Combining Instruction

The sentence combining instruction students received was designed to improve students' sentence-construction skills and promote use of these skills when writing and revising. As was done in Saddler and Graham (2005), instruction was delivered in five units. The first unit, 1 week in duration, focused on combining smaller related sentences into a more complex compound sentence using the connectors *and*, *but*,

because, and *so* (e.g., “The cake is delicious” and “The cake is chocolate” combined to “The cake is chocolate, and it is delicious”). Starting with compound sentences, using these conjunctions allowed us to begin instruction with a fairly easy sentence skill. The next unit of sentence combining, 2 weeks in duration, focused on embedding adjectives and adverbs from one sentence into the another (e.g., “They play football” and “They play very rough and quickly” combined to “They play football very rough and quickly”). The third and fourth sentence combining units, 2 weeks in duration, focused on complex sentences with embedded adverbial and adjectival clause (e.g., “The teacher stopped to talk about their salary” and “The principal came to the teachers’ room” combined to “The teachers stopped to talk about their salary when the principal came into their room”). The final unit, 1 week in duration, extended the creation of complex sentences by embedding adjectives, adverbs, adverbial clauses, and adjectival clauses (e.g., “Omer bought a shirt,” “Omer is friends with Abdullah,” “Omer tried on many t-shirts” were combined to form “Omer, who is a friend of Abdullah, bought a t-shirt”).

During the first lesson each week, instruction was scaffolded, so teachers explained and modeled while thinking aloud how to combine smaller kernel sentences into a specific kind of more complex sentence. For example, in the first unit, teachers demonstrated how two related simple sentences could be combined into a compound sentence. Next, teachers continued to create the same kinds of sentences using sentence combining, but increasingly drew upon assistance from students. Then, students independently practiced applying the targeted sentence combining procedure, with the teacher providing assistances as needed. This was followed by additional peer-assisted independent practice in applying the targeted sentence combining procedure. With this approach, each student alternatively acted as the coach to another child who was tasked with applying the sentence combining procedure. To guide the peer-assistance process, the coach was provided with a set of cards to direct the other student to (a) read the sentences to be combined out loud, (b) decide the best way to combine the sentences, (c) write the answer. and (d) read the new combined sentence. If the combined sentence was grammatically correct, the coach reinforced the peer by saying “good job.” If the sentence was not correct, the coach provided suggestions on how to fix it. If needed, the coach called on the teacher for assistance. As teachers modeled and students combined sentences individually or in pairs, the combined sentence was discussed and evaluated.

In the second weekly lesson, the student pairs worked together to revise a paragraph that included a series of related kernel sentences that could be combined using the sentence combining skills they were taught in that and the preceding unit. No clues were provided as to how the sentences should be combined, providing students with choice when deciding how to revise this paragraph. One student from the pair read the revised paragraph aloud, and the teacher and students discussed the rhetorical effect of the revisions made. The student pair was then asked to revise the sentences in this paragraph again to make it even better. These revising activities provided students with opportunities to apply learned sentence combining procedures in a writing context.

In the third lesson, pairs of students again worked together to write and revise a short story using the sentence combining skills taught in that unit and preceding

units. To facilitate this activity, students were provided with a paper planning facilitator containing three columns. The first column had two characters to be included in the story. The second column described two different settings where students could situate their story. The third column provided two possible topics for the story. Students were also provided with five kernel sentences to use when ending their story. The process of writing and revising a story where they were asked to use the sentence combining procedures learned provided students with an opportunity to apply these skills as they composed and revised text.

All lessons began with the teacher establishing the purpose of the lesson. At the end of each lesson, this purpose was reviewed. In addition, students were asked to apply the targeted sentence combining procedures at other times during the day. At the start of the next lesson, they were asked to share instances of when they applied this procedure and record how often this occurred.

The sentences combined by teachers and students in this study were taken from popular books for young children in Turkey. Sentences in selected passages in these books were deconstructed into simple, kernel sentences that could be combined into the compound and complex sentences taught in this investigation. The deconstructed passages had a first to second grade readability according to the Fry formula.

BAU Control Condition

Teachers in the control condition continued with their typical writing instruction during the experiment. Writing instruction was based on a textbook series that the school used to teach the language arts. The textbook series mostly involved a product-based approach to teaching writing, where students were asked to read a short piece of text about a topic, commonly accompanied by a picture, and then directed to write a particular response to the material read. For example, in one such activity, the textbook presents a short text of 81 words for students to read, describing two cousins, one who which has special needs. After reading the text, students are asked to write a short response: How can we make life easier for persons with special needs? Our informal observations of teachers in the control condition suggest that they rarely encouraged students to plan, revise, or edit such responses or taught strategies for doing so. Across Grades 2 to 4, the textbook series used by control teachers asked students to engage in a variety of these kinds of product-based activities including writing poems, letters, diary entries, stories, descriptions, and informative directions to provide a few examples.

Fidelity

Sentence Combining Treatment

Eleven observations were conducted to determine if treatment teachers implemented sentence combining instruction with fidelity. These observations were

conducted by the first author. Each treatment teacher was observed four times, except for the fourth-grade teacher who was observed three times. Each observation determined if the lesson was implemented as intended (i.e., the prescribed steps of the treatment were applied in each lesson), the teacher was prepared, the lesson was not interrupted or adversely impacted by classroom events, and the teacher taught only the skills targeted in the lesson. Treatment fidelity was strong, as these criteria were met 98% of the time across all observations (94% to 100% of the time depending on the teacher).

While treatment teachers were taught how to deliver the same sentence and sentence combining instruction to second- through fourth-grade teachers, our observations of these teachers indicated that they adjusted their teaching based on the age of their students (which was emphasized during the professional development they received). For instance, teachers of younger students provided more explanations when explaining a concept than teachers of older students. In contrast, older students were provided with less repetition or practice when learning how to combine sentences.

BAU Control Condition

Eleven observations were also conducted to determine if teachers in the business-as-usual control condition were implementing writing instruction as prescribed in the prescribed language arts text and activity books. Again, each control teacher was observed four times, except for the fourth-grade teacher who was observed three times. The same criteria as described above were used to judge fidelity. These criteria were met 91% of the time (83% to 100% of the time depending on the teacher).

Reliability of Observations

To establish reliability of the observations of treatment and business-as-usual control teachers, a graduate student who was familiar with instructional procedures for both groups independently observed eight lessons (an equal number for each condition). The students' observations were identical with the first authors 94% of the time.

Measures for Assessing the Impact of Sentence Combining

Writing Task

Prior to randomly assigning participating teachers to the treatment or control conditions, students in the six classrooms were asked to write an opinion essay. Immediately following the end of the sentence combining treatment, students in all classrooms were asked to write a second opinion essay. The pretest and posttest writing topics were identical, except that they were written for different audiences. At pretest, students were asked to identify something they wanted their teacher to do and

then write a paper that would convince the teacher to do this. At posttest, the same basic prompt was applied, except now students were asked to identify something they wanted their classmates to do and write a paper convincing them to do so. To be sure students understood the demands of the writing task, they were provided with an example of an opinion essay before the pretest, and they discussed the purpose and structure of such writing.

Writing Measures

Students' pre- and posttest essays were scored using a Turkish adaptation (see Özkara, 2007) adaptation of the 6 + 1 Analytic Writing and Assessment Scale developed by Northwest Regional Education Laboratory in the USA (Bellamy, 2000). This scale assessed seven aspects of writing: ideation, organization, voice, word choice, sentence fluency, conventions, and presentation. Each of these seven aspects was scored using a one- to five-point scale. For scores of 1, 3, and 5, a written description of criterion for each score was provided to raters. Higher scores represented better writing performance. For example, a score of 5 for organization indicated that the paper was clearly organized, and the reader could easily understand it; a score of 3 indicated the paper was not fully organized but the reader could mostly understand it; a score of 1 indicated that the paper was poorly organized making it difficult for the reader to understand it. Similarly, a score of 5 for conventions indicated that there were few if any spelling, grammar, or usage errors; a score of 3 indicated that spelling, grammar, and usage was mostly used correctly, but errors did occur throughout the text; a score of 1 indicated that spelling, grammar, and usage errors occurred very frequently.

The score for sentence fluency on the adapted 6 + 1 measure provided a proximal assessment for the effects of the sentence combining treatment. A score of 5 indicated that the students used different types of sentences that were fully constructed when writing; a score of 3 indicated the student used some different types of sentences and most of these were complete; a score of 1 indicated the student mainly used a single type of sentence and many of these were incomplete.

The papers students wrote at pre- and posttest also provided two more distal measures of writing performance. One was a measure of writing quality and included the average of the scores from the 6 + 1 assessment for ideation, organization, voice, word choice, conventions, and presentation. We did not include sentence fluency as part of this assessment because it more directly assessed the effects of the sentence combining treatment. A factor analysis of the pre-test data for the writing quality measure indicated that the instrument was a single factor measure (eigenvalue = 3.13; all six aspects of writing loading at 0.63 or higher) with acceptable reliability (coefficient alpha = 0.83). Likewise, a second factor analysis with posttest data yielded a single factor with an eigenvalue greater than 1.0 (eigenvalue = 3.96; all six aspects of writing loading at 0.73 or higher) with acceptable reliability (coefficient alpha = 0.90).

The second distal measure collected from students' pretest and posttest papers involved writing output. The total number of words, regardless of correct spelling or grammar, students included in their writing was counted.

Reliability of Scoring

Pretest and posttest papers were scored once the study had ended, and identifying information was removed from them before scoring. Each paper was independently scored by two raters. Pearson product moment correlations between the scores assigned by the two raters were 0.88, 0.96, and 0.99 for sentence fluency, writing quality, and number of words, respectively.

Procedures

All students in the six participating classrooms were administered the pretest by their teachers. Teachers were then randomly assigned to the sentence combining condition or the BAU control. Teachers assigned to the sentence combining condition then received 16 h of professional development from the first author. Once the professional development was completed, treatment teachers delivered 2 weeks of sentence instruction (three 2-h sessions a week), followed by 8 weeks of sentence combining instruction (three 2-h sessions a week). Each week during instruction, treatment teachers were provided with 1 h of coaching from the first author. Treatment fidelity observations for the sentence combining and BAU control conditions were also conducted during the 10-week treatment period. The week after sentence combining instruction ended, students in all six classes (treatment and control) completed the posttest which was administered by their teacher. Two weeks following the posttest, teachers in the BAU control condition were provided with a professional development session on how to apply the sentence combining treatment.

Data Analysis

To investigate the effects of the sentence combining treatment, we estimated separate random effects multilevel models for each of the three writing outcome measures: (a) sentence fluency, (b) writing quality, and (c) writing output (Snijders & Bosker, 2011). The dataset was complete without any missing participants data for pretest (before sentence combining instruction) or posttest (after sentence combining instruction) among the three writing measures. Because the 171 students were nested within 6 classes, we examined Intraclass Correlations Coefficients (ICCs) to estimate the extent of potential cluster effects. Classroom ICCs ranged from 0.12 to 0.34 ($M=0.22$, $SD=0.09$). Therefore, each random effects, two-level model, accounted for the nested data structure with students at Level 1 ($N=171$) and teachers at Level 2 ($N=6$).

Initially, we analyzed unconditional models for the three outcomes (Model 0). Then, we estimated multilevel models with pretest scores entered as a fixed-effect covariate at Level 1 to control for initial writing performance before students received the treatment (Maxwell, 1998). All pretest raw scores were grand-mean centered before we entered them into the models. The multilevel models also included dummy and contrast variables to examine the effects of

the sentence combining treatment for the full sample and by grade. In a build-up approach (Hox, 2010), we entered a dummy variable coded with the sentence combining group as 1 versus the BAU control condition as 0 (Model 1). Next, we entered contrast variables into the models to compare student writing performance in Grades 3 and Grades 4 (each coded as 1 in the respective comparison) to the referent group, Grade 2 (coded as 0 in Model 2). Finally, we ran models estimating interaction effects of the treatments by grade ($TxGroup_{ij} \times Grade_j$) as well as main effects of the experimental conditions and contrasts of grade level performance (Model 3). We used the “mixed” command to estimate all models in Stata 14.2 (StataCorp, 2015). The conditional multilevel models estimated sentence combining effectiveness and controlling for pretest assumed the following form:

$$\begin{aligned}
 Posttest_{ij} = & \gamma_{00} + \gamma_{10}Pretest_{ij} + \gamma_{20}TxGroup_{ij} + \gamma_{01}Grade3_j + \gamma_{02}Grade4_j \\
 & + \gamma_{21}TxGroup_{ij} \times Grade3_j + \gamma_{22}TxGroup_{ij} \times Grade4_j + u_{0j} + e_{ij}
 \end{aligned}$$

After fitting the multilevel models, we estimated Hedge’s *g* standardized mean difference effect sizes with cluster adjustments to determine the magnitude of the treatment effects between conditions (sentence combining instruction vs. BAU control) and the effects of conditions across grade levels (i.e., Grades 2, 3, and 4; Borenstein et al., 2009).

Table 1 Means and standard deviations by study condition and grade (*N* = 171)

Measure by sample	Sentence combining						BAU control			
	Pretest			Posttest			Pretest		Posttest	
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Sentence fluency										
Grade 2	30	1.07	0.37	3.07	1.23	28	1.64	0.95	1.71	0.98
Grade 3	25	2.20	1.00	3.40	1.41	23	1.96	1.02	2.23	1.00
Grade 4	33	2.58	1.20	4.33	0.96	32	2.19	1.00	3.38	1.39
Total	88	1.95	1.13	3.64	1.31	83	1.94	1.00	2.49	1.36
Writing output										
Grade 2	30	40.47	17.32	59.80	17.31	28	40.32	19.73	38.11	14.31
Grade 3	25	69.16	30.50	90.28	38.84	23	54.30	22.91	60.87	23.23
Grade 4	33	70.73	33.35	88.48	25.65	32	69.13	30.50	65.59	18.85
Total	88	59.97	31.07	79.22	30.92	83	55.30	27.78	55.01	22.30
Writing quality										
Grade 2	30	1.63	0.57	2.98	0.97	28	1.81	0.72	2.25	0.79
Grade 3	25	2.16	0.57	3.63	1.03	23	1.91	0.54	2.68	0.62
Grade 4	33	2.56	1.15	4.03	0.55	32	2.30	0.64	3.30	0.81
Total	88	2.13	0.92	3.56	0.96	83	2.03	0.67	2.77	0.87

N = 171 students nested within 6 classes (2 per grade)

Results

Descriptive statistics for pretest and posttest across the three writing outcomes (sentences, writing output, and writing quality) by experimental condition and grade are reported in Table 1. Estimates for skewness and kurtosis across all writing measures were less than 1.00 for the total sample of participants. Inspection of histograms suggested a slight right skew in the sentence fluency measure at pretest, with more students scoring on the lower end of the scale at pretest. All other outcome data was normally distributed. Table 2 reports pairwise correlation coefficients for pretest and posttest measures. Correlations significant at p less than 0.05 ranged from 0.16 to 0.83 on the posttest measures.

Does Sentence Combining Instruction Improve Students' Sentence Skills? (RQ1)

Across the full sample ($N=171$), the unconditional model for the sentence fluency outcome measure indicated substantial variability among classes ($\tau_{00} = 0.68$) and students (residual $\sigma^2 = 1.38$), with an ICC estimated as 0.33. Table 3 reports estimates from the series of separate random effects multilevel models run for the sentence fluency outcome measure. When controlling for pretest performance, results in Model 1a indicated statistically significant differences between the sentence combining treatment and the BAU control group ($\gamma_{20} = 1.16$; $p=0.02$; 95% CI [0.22, 2.10]). An effect size of 0.85 indicated students, on average, performed higher at posttest after receiving the sentence combining instruction (see Table 4). Effect sizes within grade-level samples comparing sentence combining instruction and the BAU control groups ranged from 0.80 to 1.19 for the sentence fluency measure at posttest (see Table 4). Model 2a in Table 3 also indicated statistically significant effects of grade, such that students in Grade 4 performed higher on the sentence fluency measure posttest when compared students in Grade 2 ($\gamma_{02} = 1.39$; $p<0.001$; 95% CI [0.94, 1.84]), regardless of the experimental condition. In contrast, results revealed no statistically significant differences between student performance Grades 2 and 3 on the sentence fluency measure. Finally, results for the interaction model (Model 3a) indicated main effects for the sentence combining treatment over the BAU control ($\gamma_{20} = 1.40$; $p<0.001$; 95% CI [0.80, 2.00]), regardless of grade. Moreover, results indicated

Table 2 Correlations of writing measures at pretest and posttest

Measure by time point	1	2	3	4	5	6
1. Sentence fluency pretest	—					
2. Sentence fluency posttest	0.21**	—				
3. Writing output pretest	0.09	0.13	—			
4. Writing output posttest	0.29***	0.65***	0.16*	—		
5. Writing quality pretest	0.73***	0.21**	0.13	0.28***	—	
6. Writing quality posttest	0.24**	0.83***	0.14	0.68***	0.21**	—

* $p<0.05$; ** $p<0.01$; *** $p<0.001$

Table 3 Multilevel models for sentence combining treatment on sentence fluency measure

	Model 1a: treatment only			Model 2a: treatment and grade			Model 3a: interaction model					
	Est	SE	<i>z</i>	<i>p</i>	Est	SE	<i>Z</i>	<i>p</i>	Est	SE	<i>z</i>	<i>p</i>
Fixed effects												
Intercept, γ_{00}	2.45	0.34	7.19	<0.001	1.85	0.19	9.97	<0.001	1.74	0.22	7.92	<0.001
Pretest, γ_{10}	0.11	0.09	1.22	0.22	0.06	0.09	0.67	0.50	0.08	0.09	0.89	0.37
Treatment, γ_{20}	1.16	0.48	2.41	0.02	1.15	0.18	6.52	<0.001	1.40	0.31	4.55	<0.001
Grade3, γ_{01}					0.37	0.24	1.57	0.12	0.48	0.33	1.46	0.14
Grade4, γ_{02}					1.39	0.23	6.07	<0.001	1.62	0.30	5.34	<0.001
Tx × Grade3, γ_{21}									-0.24	0.46	-0.52	0.60
Tx × Grade4, γ_{22}									-0.47	0.43	-1.11	0.27
Random effects												
Classroom (τ_{00})	0.30				<0.001							
Student/residual (σ^2)	1.38				1.34							

Est. estimate, *SE* standard error, *Tx* treatment for sentence combining, *N* 171 students nested within 6 classes (2 per grade)

Table 4 Effect sizes for sentence combining on three writing outcomes for the full sample ($N=171$) and by grade

	Writing measure ES_g		
	Sentence fluency	Writing output	Writing quality
Grade 2	1.19	1.34	0.81
Grade 3	0.94	0.89	1.08
Grade 4	0.80	1.00	1.05
Full sample	0.85	0.88	0.84

ES_g Hedge's g standardized mean difference effect sizes with cluster adjustments; $N=171$ students nested within 6 teachers' classes

students performed higher in Grade 4 versus Grade 2 ($\gamma_{02} = 1.62$; $p < 0.001$; 95% CI [1.02, 2.21]), regardless of experimental condition. However, no interactions between the treatment and grade levels were statistically significant.

Does Sentence Combining Instruction Increase the Quality and Output of Students' Opinion Essays? (RQ2)

For writing output, the unconditional model revealed substantial variability among classes ($\tau_{00} = 301.16$) and students (residual $\sigma^2 = 571.87$). The ICC for writing output was estimated as 0.34. Table 5 reports results for the series of separate random effects multilevel models run for the writing output measure at posttest. Model 1b revealed statistically significant differences when comparing performance of students who received the sentence combining treatment versus BAU control group ($\gamma_{20} = 24.92$; $p = 0.03$; 95% CI [2.89, 46.95]). As such, students, on average, performed 0.88 standard deviations higher at posttest in the sentence combining group over BAU control students (see Table 4). In addition, effect sizes within grade-level samples comparing sentence combining and the BAU control group ranged from 0.89 to 1.34 (see Table 4), suggesting that a large effect for the sentence combining treatment was observed within each grade. Statistically significant effects of grade were also measured on writing output at posttest. These results indicated that students in Grades 3 ($\gamma_{01} = 28.36$; $p < 0.001$; 95% CI [18.94, 37.79]) and Grade 4 ($\gamma_{02} = 30.29$; $p < 0.001$; 95% CI [21.10, 39.49]) wrote more words than students in Grade 2, regardless of experimental condition. Finally, interaction in Model 3b indicated main effects of writing output for sentence combining over BAU control ($\gamma_{20} = 21.70$; $p < 0.001$; 95% CI [9.66, 33.75]). Furthermore, students wrote more words at posttests in Grade 3 compared to Grade 2 ($\gamma_{01} = 23.91$; $p < 0.001$; 95% CI [10.87, 36.94]) and Grade 4 compared to Grade 2 ($\gamma_{02} = 29.85$; $p < 0.001$; 95% CI [17.37, 42.32]), regardless of experimental condition. No interactions estimated between the treatment and grade levels were statistically significant for writing output.

The unconditional model for writing quality revealed substantial variability among classes ($\tau_{00} = 0.32$) and students (residual $\sigma^2 = 0.65$), with an ICC estimated as 0.33. Table 6 reports results for the series of separate random effects multilevel models for writing quality outcomes at posttest. When comparing writing quality for students who received sentence combining versus BAU control group, Model 1c revealed statistically significant differences between the experimental conditions ($\gamma_{20} = 0.80$; $p = 0.02$;

Table 5 Multilevel models for sentence combining treatment on writing output measure

	Model 1b: treatment only			Model 2b: treatment and grade			Model 3b: interaction model					
	Est	SE	<i>z</i>	<i>p</i>	Est	SE	<i>z</i>	<i>p</i>	Est	SE	<i>z</i>	<i>p</i>
Fixed effects												
Intercept, γ_{00}	54.69	7.96	6.87	<0.001	35.30	3.84	9.19	<0.001	36.68	4.58	8.01	<0.001
Pretest, γ_{10}	-0.06	0.07	-0.81	0.42	-0.07	0.07	-1.08	0.28	-0.08	0.07	-1.20	0.23
Treatment, γ_{20}	24.92	11.24	2.22	0.03	24.67	3.61	6.84	<0.001	21.70	6.15	3.53	<0.001
Grade3, γ_{01}					28.36	4.81	5.90	<0.001	23.91	6.65	3.59	<0.001
Grade4, γ_{02}					30.29	4.69	6.46	<0.001	29.84	6.36	4.69	<0.001
Tx × Grade3, γ_{21}									8.92	9.19	0.97	0.33
Tx × Grade4, γ_{22}									1.32	8.45	0.16	0.88
Random effects												
Classroom (τ_{00})	169.05				<0.001				<0.001			
Student/Residual (σ^2)	567.52				550.64				546.20			

Est. estimate, *SE* standard error, *Tx* treatment for sentence combining, *N*=171 students nested within 6 classes (2 per grade)

Table 6 Multilevel models for sentence combining treatment on writing quality measure

	Model 1c: treatment only			Model 2c: treatment and grade			Model 3c: interaction model					
	Est	SE	z	p	Est	SE	z	p	Est	SE	z	p
Fixed effects												
Intercept, γ_{00}	2.75	0.24	11.33	<0.001	2.23	0.13	17.59	<0.001	2.26	0.15	14.87	<0.001
Pretest, γ_{10}	0.06	0.08	0.66	0.51	0.03	0.08	0.36	0.72	0.03	0.08	0.31	0.76
Treatment, γ_{20}	0.80	0.34	2.32	0.02	0.79	0.12	6.47	<0.001	0.73	0.21	3.50	<0.001
Grade3, γ_{01}					0.54	0.16	3.40	0.001	0.43	0.22	1.92	0.06
Grade4, γ_{02}					1.03	0.16	6.62	<0.001	1.03	0.21	4.91	<0.001
Tx × Grade3, γ_{21}									0.21	0.31	0.66	0.51
Tx × Grade4, γ_{22}									<0.001	0.29	<0.001	1.00
Random effects												
Classroom (τ_{00})	0.15				<0.001				<0.001			
Student/residual (σ^2)	0.65				0.63				0.63			

Est. estimate, SE standard error, Tx treatment for sentence combining; N = 171 students nested within 6 classes (2 per grade)

95% CI [0.12, 1.47]). An effect size of 0.84 indicated students performed higher at posttest after receiving sentence combining instruction (see Table 4). Within grade-level effect sizes ranged from 0.81 to 1.08 for writing quality at posttest (see Table 4), suggesting a large effect for sentence combining at each grade.

Statistically significant effects of grade were also revealed in Model 2c (see Table 6). These results indicated that students in Grade 3 ($\gamma_{01} = 0.54$; $p = 0.001$; 95% CI [0.23, 0.84]) and Grade 4 ($\gamma_{02} = 1.03$; $p < 0.001$; 95% CI [0.72, 1.33]) wrote higher quality essays at posttest when compared to students in Grade 2, regardless of experimental condition. Finally, examination of treatment by grade interactions (see Model 3b in Table 6) revealed main effects of writing quality for the sentence combining treatment over control ($\gamma_{20} = 0.73$; $p < 0.001$; 95% CI [0.32, 1.14]). In addition, students wrote higher quality essays in Grade 4 compared to Grade 2 ($\gamma_{02} = 1.03$; $p < 0.001$; 95% CI [0.62, 1.44]), regardless of experimental condition. None of the interactions between treatment and grade level were statistically significant for writing quality. Moreover, main effects comparing performance of students in Grade 3 to Grade 2, regardless of experimental condition, were not significant in the interaction model.

Discussion

Sentence Combining Improved Turkish Students' Sentence Skills When Writing

We predicted that sentence combining instruction would enhance students' sentence fluency, which was assessed as the use of different and fully constructed sentences when writing opinion essays. Students who received such instruction practiced combining simple sentences into more sophisticated compound and complex sentences and then applied these newly learned sentence skills when writing and revising text. As anticipated, sentence fluency scores for students in the treatment condition statistically exceeded the scores of students in the BAU control condition. The effect size for sentence fluency was 0.85, indicating that sentence combining had a large impact on improving sentence skills in the context of writing.

This finding is important for three reasons. One, it replicated earlier studies showing that sentence combining instruction can improve students' sentence construction skills (Andrews et al., 2006; Limpo & Alves, 2013; Saddler & Graham, 2005). Two, it provided additional support for using a peer-assisted approach to sentence combining, going beyond Saddler and Graham (2005) demonstration that such a method delivered to small groups of students improved the sentence skills of more and less capable fourth-grade writers, by showing this approach improved sentence skills when provided through whole-group instruction to typically developing students in Grades 4 as well as grades 2 and 3. Three, this study demonstrated that sentence combining instruction is effective outside of US and European contexts. This supports the contention that sentence combining instruction can be applied broadly, at least to countries that use an alphabetic writing system as does Turkey.

Additional research is needed to determine if sentence combining instruction has similar positive effects on the sentence skills of students in other countries outside of Turkey, the USA, and Europe, including countries that use logographic and syllabic

writing systems. Future research with Turkish students and those from other countries needs to examine the effects of sentence combining instruction with middle and high school students as well as with a variety of different types of students, including students who find writing challenging, students with migration backgrounds, and students with different socio-economic circumstances. We further encourage researchers to test the effectiveness of different types of sentence combining programs, including ones that involve peer-assisted learning and ones that do not.

Sentence Combining Improved Writing Quality and Output of Turkish Students

We further expected that the effects of sentence combining instruction would have positive impacts on other aspects of students' writing beyond sentence construction. As students' sentence construction skills become more sophisticated and habitual as a result of sentence combining instruction, this should free cognitive resources that students can apply to other important writing processes such as conceptualization, ideation, and reconceptualization (Graham, 2018), resulting in longer and better text. As predicted, the length and quality of the opinion essays written by students receiving sentence combining instruction statistically exceeded that of students in the BAU control condition. Effect sizes for writing quality and length of opinion essays were 0.84 and 0.88, respectively, indicating that sentence combining instruction had a large impact on these two aspects of writing.

The positive effects of sentence combining on the quality of students' writing in this study were consistent with the outcomes from studies conducted in the USA and Europe (Andrews et al., 2006; Hillocks, 1986; Graham & Perin, 2007; Koster et al., 2015; Limpo & Alves, 2013; Saddler & Graham, 2005; Walters et al., 2021). The finding that sentence construction enhanced the length of students' opinion essays was consistent with Limpo and Alves (2013), but inconsistent with the finding for writing length reported by Saddler and Graham (2005). This latter finding was unexpected because the sentence combining procedures applied in the present study were based on the ones used by Saddler and Graham (2005). There were a number of possible differences between this and the Saddler and Graham (2005) investigation that may be responsible for this difference. This included a focus on different grade levels (Grades 2 to 4 vs just Grade 4), type of student (typically developing writers vs more and less capable writers), writing tasks (opinion essays vs story writing), and different cultural and educational contexts (Turkey vs the USA). Researchers need to more consistently collect data on the effects of sentence combining instruction on the length of students' essays so a clearer picture of the effects of such instruction on this writing outcome can be determined.

The findings from this study and prior investigations (Andrews et al., 2006; Hillocks, 1986; Graham & Perin, 2007; Koster et al., 2015; Limpo & Alves, 2013; Saddler & Graham, 2005; Walters et al., 2021) demonstrating that sentence combining improved not only sentence skills, but other important aspects of writing provide support for the assumed importance of translations skills to writing as depicted in the WWC model (Graham, 2018) and other models of writing (e.g., Hayes, 1996). Given the frequency with which sentence instruction has been shown to improve both sentence skills and the quality of students' writing, we encourage researchers to devote more attention to

determine how sentence skills develop longitudinally, the effort and cognitive resources required by sentence construction at different levels of writing development, and the interplay between sentence construction and other writing production processes.

The effects of sentence combining in this study with Turkish students' produced outcomes for writing quality and writing output that typically exceeded those for sentence combining studies in USA and European studies (see Andrews et al., 2006; Hill-ocks, 1986; Graham & Perin, 2007; Koster et al., 2015; Limpo & Alves, 2013; Walters et al., 2021). It is not certain if this is also the case for sentence skills due to differences in how such skills were tested in this investigation (in context) and prior ones (often as isolated skills). The outcomes from this study, when compared to prior sentence combining investigations, provide tentative support for the assumption in the WWC model (Graham, 2018) that the effectiveness of writing treatments likely varies from one country to another (see also Graham, 2021). However, this theoretical proposition needs to be scientifically tested more directly by comparing the effects of the same writing treatments across countries that differ in meaningfully ways including writing systems as well as cultural, social, historical, institutional, and political determinants.

The Effects of Sentence Combining Instructions Were Constant Across Grades

While older students in this study generally evidenced higher scores for sentence fluency, length of essays, and quality of writing than younger ones, there were no statistically detectable interaction effects for grade (Grades 2 to 4) and instructional condition (sentence combining instruction vs BAU control) for these three outcomes. Consequently, the sentence combining program was not more or less effective for students in the three grades included in this investigation. This may have been a consequence of the sentence instruction provided in the first two weeks of the program which was designed to ensure that all students had the background skills needed to benefit from the provided sentence combining instruction. It could also be due to adjustments teachers made when teaching the program to their students, including teachers of younger students providing more explanation and repeated practice than teachers of older students. In any event, we hope that future studies with students in multiple grades will design sentence combining instruction that becomes increasingly sophisticated as students move from one grade to the next. This has the potential to result in sentence combining programs that are even more effective than the one tested in this investigation.

Limitations, Implications, and Conclusions

As with all studies, the current investigation has limitations that must be considered when interpreting the obtained findings. First, while the study involved 171 students, it included only six classrooms, with two classrooms at each grade level. It is possible that the calculated ICCs may reflect writing differences across grade levels as well as among the classroom clusters of students. Nonetheless, because students were randomly assigned to the sentence combining or BAU control at the teacher level, our data analyses accounted for the multilevel structure of students nested within classrooms.

Second, outcome measures were based on students' opinion writing. We selected this form of writing because Turkish students indicated in prior research that they preferred it to other forms of writing (Seban, 2016; Tavşanlı, 2018). However, there is no guarantee that our findings will generalize to other types of writing. This assumption must be tested empirically. Third, the current study was conducted in a single country. While we expect that sentence combining instruction is a highly portable and effective approach for teaching sentence construction skills, the application of the program tested here may yield smaller or even larger effects depending on the country in which it is applied (Graham, 2018).

Third, unlike the Saddler and Graham (2005) who compared sentence combining to grammar instruction, our control condition was BAU. This had the advantage of providing an assessment of sentence combining instruction when compared to typical writing practices in Turkey. We hope that subsequent replications will compare sentence combining instruction to both grammar instruction and BAU, providing multiple comparisons for judging the impact of sentence combining instruction.

In summary, the findings from the current study were replicated and extended previous research (Saddler and Graham, 2005) showing that peer-assisted sentence combining instruction improves the sentence-construction skills of young developing writers and that such instruction can enhance these writing output and quality. To date, there are few scientifically validated practices for teaching sentence construction skills. As this paper and other investigations have demonstrated, sentence combining provides teachers with a tool to help students become better writers. As with any evidence-based practice, teachers who apply a sentence combining approach need to carefully monitor its success and make needed adjustments when it is not effective.

References

- Andrews, R., Torgerson, C., Bevertson, S., Freeman, A., Locke, T., Low, G., Robinson, A., & Zhu, D. (2006). The effects of grammar teaching on writing development. *British Educational Research Journal*, 32, 39–55. <https://doi.org/10.1080/01411920500401997>
- Aşıkcan, M., & Pilten, G. (2016). The evaluation of classroom teachers' written expression implementations in terms of process based writing model. *Turkish Studies*, 11(3), 255–276.
- Babayiğit, S., & Stainthorp, R. (2010). Component processes of early reading, spelling, and narrative writing skills in Turkish: A longitudinal study. *Reading and Writing*, 23, 539–568. <https://doi.org/10.1007/s11145-009-9173-y>
- Bellamy, P. C. (2000). *Research on writing with the 6 + 1 traits*. Portland, OR: Northwest Regional Educational Laboratory.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.
- Gein, J. van de (1991). *The sense of sentences. A study into the effects of grammar instruction upon junior writing*. Doctoral dissertation. Utrecht: Utrecht University.
- Graham, S. (2006). Writing. In P. Alexander & P. Winne (Eds.), *Handbook of Educational Psychology* (pp. 457–478). Mahwah, NJ: Erlbaum.
- Graham, S. (2018). A revised writer (s)-within-community model of writing. *Educational Psychologist*, 53(4), 258–279. <https://doi.org/10.1080/00461520.2018.1481406>
- Graham, S. (2021). Executive Control and the Writer(s)-Within-Community Model. In T. Limpo & T. Olive (Eds.), *Executive functions and writing* (pp. 38–76). Oxford, UK: Oxford Press.

- Graham, S., & Harris, K. R. (2014). Conducting high quality writing intervention research: Twelve recommendations. *Journal of Writing Research*, 6, 89–123.
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99, 445–476. <https://doi.org/10.1037/0022-0663.99.3.445>
- Graham, S., Kiuahara, S., McKeown, D., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology*, 104, 879–896.
- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1–27). Erlbaum.
- Hillocks, G. (1986). *Research on written composition*. Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills. (ERIC Document Reproduction Service No. ED265552)
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). Routledge/Taylor & Francis Group.
- Hsiang, T., & Graham, S. (2016). Teaching writing in grades 4 to 6 in urban schools in the Greater China region. *Reading & Writing: An International Journal*, 29, 869–902. <https://doi.org/10.1007/s11145-015-9597-5>
- Hsiang, T., Graham, S., & Wong, P. (2018). Teaching writing in grades 7–9 in urban schools in the Greater China Region. *Reading Research Quarterly*, 53, 473–507.
- Kamal, K. (2017). Middle East: Education in Turkey. *World Education News and Reviews*. <https://wenr.wes.org/2017/04/education-in-turkey>
- Koster, M., Tribushinina, E., de Jong, P. F., & van den Bergh, H. (2015). Teaching children to write: A meta-analysis of writing intervention research. *Journal of Writing Research*, 7(2), 249–274.
- 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>
- Maxwell, S. E. (1998). Longitudinal designs in randomized group comparisons: When will intermediate observations increase statistical power? *Psychological Methods*, 3(3), 275–290.
- Özkara, Y. (2007). *6 + 1 analytic writing and evaluative model on enhancing 5th grade students narrative writing skills [Unpublished doctoral dissertation]*. Gazi University.
- Saddler, B. (2012). *Teacher's guide to effective sentence writing*. Guilford.
- 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>
- Seban, D. (2016). The role of gender and grade level on topic selection in writing. *Mediterranean Journal of Humanities*, 6(1), 275–285.
- Snijders, T. A. B., & Bosker, R. J. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. SAGE.
- StataCorp. (2015). *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.
- Tavşanlı, Ö. F. (2017). Evaluation of the Instructional Program in Turkey Based on the Process-Based Writing Approach. *International Journal of Language Academy*, 5(2), 79–97.
- Tavşanlı, Ö. F. (2018). What do primary school students write about? Investigation of text types, subjects and content preferences in the written works of primary school students. *Journal of Mother Tongue Education*, 6(1), 32–47.
- Tavşanlı, Ö. F., & Kara, Ü. E. (2022). Developing the draft papers of primary school fourth-grade students. *Through Review and Feedback-Based Editorial Study*. HAYEF: Journal.
- Türkben, T. (2021). Examination of classroom writing practices in the context of process-based approaches to teaching writing. *Journal of Language & Linguistic Studies*, 17, 619–644.
- Walters, K., Dockrell, J., & Connelly, V. (2021). A sentence-combining intervention for struggling writers: Response to intervention. *Reading & Writing: An Interdisciplinary Journal*, 34(7), 1825–1850.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.