PEDAGOGISCHE STUDIËN
TIJDSCHRIFT VOOR ONDERWIJSKUNDE EN OPVOEDKUNDE

# Text structure instruction in primary education: Effects on reading, summarization, and writing

**S.T.M. Bogaerds-Hazenberg, J. Evers-Vermeul, & H.H. van den Bergh**

**Summary** Although knowledge of informational text structures can promote text comprehension, this topic receives little attention in the Dutch primary school curriculum. 201 Dutch students in grades 4-6 participated with their teachers ($n$ = 10) in this quasi-experimental study with a switching-panels replication design. Students either first followed a text structure intervention (TOP) and then went back to business-as-usual, or the other way around. During the intervention, teachers taught their students about the characteristics of four informational text structures, and how to use structure-specific graphic organizers to organize main ideas for each structure. In addition, several writing tasks related to the different text structures were included. At three measurement occasions, students completed text structure tests, reading comprehension tests, summarization tasks, and writing tasks. Only the fourth graders in one iteration of the intervention showed immediate effects over and above the effect of business-as-usual lessons on the text structure test ($d$ = 0.50), the reading comprehension test ($d$ = 0.53), the summarization task ($d$ = 0.48). In both iterations of the intervention, an immediate effect was found on writing ($d$ = 0.33 and $d$ = 0.39). These findings are discussed in the light of test-related issues and implementation fidelity data.

**Keywords** text structure instruction, reading comprehension, primary education, implementation fidelity

**Artikelgeschiedenis**
Ontvangen: 21 maart 2023
Ontvangen in gereviseerde vorm: 12 december 2023
**Geaccepteerd**:
19 december 2023
Online: 4 maart 2024
**Contactpersoon**
Suzanne Bogaerds-Hazenberg, suzanne.bogaerds-hazenberg@ru.nl

**Financiering onderzoek**
Funded by Netherlands Organization for Scientific Research (NWO), Grant Number: 406.16.519
**Belangen**
De auteurs hebben geen belangen te vermelden.


https://doi.org/10.59302/ps.v101i1.18783
2024 (101) 30-59

## 1 Introduction

Despite intensive reading instruction, a substantial number of students struggle with reading comprehension, especially when it comes to the processing of informational texts (e.g., Helder et al., 2016; Kraal et al., 2018; Rooijackers et al., 2020; Wijekumar et al., 2017). This problem also holds true for the Dutch context where many students appear to struggle with integrating information, as well as with summarizing texts (e.g., Dutch Inspectorate of Education, 2022; PPON-54: Kuhlemeier et al., 2014). These comprehension problems often become apparent on the threshold of the upper elementary grades, when the focus shifts from learning-to-read to reading-to-learn (Harlaar et al., 2017).

One reason for these comprehension problems may be the sudden increase of informational texts to be read around fourth grade. Informational texts are generally assumed to be less familiar and more complex than narrative texts. Besides a difficult vocabulary and high information density, they also have complex text structures (Coté et al., 1998), such as compare-contrast, chronology, problem-solution or cause-effect (Table 1). Students are often less familiar with such structures than with the structure of narrative texts (e.g., Hiebert & Mesmer, 2013).

A second reason might be that reading-to-learn requires advanced text comprehension skills: students need to make connections across different parts of text, identify main ideas and relate these to their prior knowledge, and evaluate what they read. This echoes the Construction-Integration Model (e.g., Kintsch, 2004), which states that, for comprehension, readers parse textual input into concepts and relationships, which they need to organize in associative networks. Ideally, readers make connections between ideas within the text, and integrate these ideas with prior knowledge, until a coherent mental representation of the text arises: the so-called situation model.

Situation model comprehension requires that readers carefully think about the main ideas and how these are linked across the text. This process can be fostered by insight in the underlying text structure. That is, a reader who recognizes the underlying text structure is like a traveler with a road map that highlights the main routes and gives insight into what is coming next. Likewise, knowledge about text structures helps readers predict upcoming information, and identify and organize main ideas and their implicit and explicit relationships more easily (e.g., Meyer & Ray, 2011; Meyer et al., 2018; Strong, 2020; Wijekumar et al., 2017; Williams et al., 2016). For instance, students who read due to in a science text should infer that a cause-effect will follow, and students who recognize a problem-solution structure will realize that after the presentation of the problem, the text will probably discuss several solutions, even if these are not explicitly highlighted with signaling words.



PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/ps.v101i1.18783

*Table 1*

Example of four text structures with signaling words and phrases in italics

| **Texts with compare-contrast structure explain how things are similar and different.** | **Texts with cause-effect structure tell how an event culminates in an outcome or effect.** |
|---|---|
| Influenza *and* COVID-19 are both contagious respiratory illnesses, but they are caused by *different* viruses. COVID-19 is caused by a coronavirus, *whereas* flu is caused by an influenza virus. The symptoms of flu and COVID-19 are *similar*: both cause symptoms like fever and shortness of breath. | Scientists still aren't sure why COVID-19 is *causing* clots. Clots might be *the result of* blood vessels' *reactions* to being invaded by the virus. The virus attacks cells via a receptor called ACE2. *When* the virus binds to these receptors in the walls of blood vessels, they become inflamed, which *can cause* clotting. |
| **Texts with chronological structure explain an order or cycle of events and/or procedures.** | **Texts with problem-solution structure explain a problem, and how it can be fixed.** |
| On *1st January 2020*, the Wuhan seafood marked was closed for inspection, as it appeared that many visitors developed a new disease. *One week later*, Chinese authorities identified a new coronavirus. *Five days later* the Chinese government shared the genetic sequence of the virus. | The COVID-19 pandemic resulted in a shortage of face masks. Also, many people were furloughed and needed new jobs. A social enterprise in India *addressed both of these problems:* former train workers started producing face masks. |



PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/ps.v101i1.18783

Text structure inference skills are positively correlated with expository text comprehension (Welie et al., 2018). Therefore, it seems promising to explicitly teach students about these text structures. Various researchers have developed intervention programs in which students learn to recognize text structures, use structure-specific main idea sentences or graphic organizers to summarize text, and make structure-specific inferences about text content (e.g., Broer et al., 2002; Williams et al., 2004), sometimes followed by structure-specific writing tasks (e.g., Strong, 2020). Text structure instruction appears effective for students of various ages to improve their text comprehension, text recall, and summarization skills (Bogaerds-Hazenberg et al., 2021; Hebert et al., 2016; Pyle et al., 2017), but several questions remain that require additional research.

First, the generalizability of findings to other populations and educational contexts is relatively unclear (Bohaty et al., 2015). The number of intervention studies outside the US is gradually increasing (e.g., China: Xu et al., 2021), but the majority of text structure research is still US based. It is important to study the effectiveness of text structure interventions in other linguistic and cultural contexts with varying reading curricula and teacher training programs (Bohaty et al., 2015; Williams, 2018), such as the Netherlands where the primary school curriculum is very implicit about informational text structures (Bogaerds-Hazenberg et al., 2022). As there are reasonable concerns on Dutch students' reading achievements, in particular with respect to higher-order

text comprehension (e.g., Dutch Inspectorate of Education, 2022; PISA-2018; Gubbels et al., 2019; van den Broek et al., 2021), this country provides an interesting context to study the effects of text structure instruction.

Second, a plea has been made that text structure research should include multiple outcome measures for reading, and examine transfer effects more carefully (Bohaty et al., 2015; Hebert et al., 2016; Williams, 2018). That is, the effects of text structure instruction on text comprehension vary greatly per type of outcome measure (Bogaerds-Hazenberg et al., 2021; Hebert et al., 2016; Pyle et al., 2017) and potential transfer effects on writing skills are much less examined. Yet, explicit text structure knowledge can also promote writing skills (Hebert et al., 2018; Strong, 2020), as it provides writers with a means to organize information on paper (Dickson, 1999). Furthermore, studies that measure delayed effects beyond the week after finishing the intervention are scarce (Hebert et al., 2016). Therefore, our study includes a broad spectrum of reading tests, while exploring effects on writing as well. The switching panels replication design (Shadish et al., 2002) of the current study allows us to replicate the intervention and test for maintenance effects 13 weeks after the intervention (see Method section).

Third, text structure research should examine powerful instructional approaches: now that we know what to teach in text structure interventions, we need to know how to successfully teach this (Hebert et al., 2016; Williams, 2018). According to Pyle and colleagues (2017), text structure instruction should ideally follow a gradual release of responsibility (GRR; e.g., Fisher & Frey, 2021). Research on reading comprehension in general suggests that GRR elements such as high-quality teacher-led instruction and modeling, followed by collaborative and individual practice matters for reading comprehension (McVee et al., 2019; Young, 2017). The meta-analysis by Bogaerds-Hazenberg et al. (2021) examined GRR elements in text structure interventions and suggested that individual practice might affect the maintenance of effects, but also highlighted how instructional approaches are often poorly described and operationalized.

For example, only a handful of text structure interventions use structured forms of collaborative learning, even though such activities can facilitate scaffolding, promote effective student interactions, and increase student motivation (Puzio & Colby, 2013). The key criteria for effective collaborative learning are (1) positive interdependence, (2) individual accountability, (3) promotive interaction, and (4) equal participation of students (Johnson & Johnson, 2017; Topping et al,. 2017), which may be promoted by offering structured collaborative learning tasks (Kagan, 2021). Therefore, the current intervention followed a GRR pattern with ample opportunities for structured collaborative learning. Our research question was as follows:



PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/ps.v101i1.18783

RQ: What are the effects of the text structure intervention program TOP on reading comprehension, text structure knowledge, summarization skills, and writing for fourth and fifth graders?

## 2 Method

### 2.1 Intervention (TOP) and business-as-usual (BAU)

The text structure intervention TOP was developed in close collaboration with four co-designing primary school teachers. Its core design principles are (1) Embed reading instruction in content-area subjects; (2) Focus on informative text structures; (3) Balance declarative, procedural, and conditional knowledge; (4) Ensure a gradual release of responsibility (see for an extensive discussion of the design principles Bogaerds-Hazenberg et al., 2019).

TOP consists of ten one-hour lessons on four text structures, subsequently on the compare-contrast, problem-solution, chronology, and cause-and-effect structure. During the intervention, lessons were completed at a rate of one or two lessons per week, while the business-as-usual group continued their existing routines.

Table 2 summarizes the content and the instructional approach of the TOP intervention. The first lesson introduced students to the concept of text structure and emphasized why it is useful to recognize text structure. This was followed by two or three lessons that focused on each structure's main components and key characteristics, the structure-specific graphic organizers, and how students could use certain reading strategies (questioning, summarizing, predicting) in the context of that structure (e.g., Which questions are useful to ask while reading a compare-contrast text?). The texts were related to various content-area subjects: eight texts were related to the biology curriculum, five to geography, and three to history as this provides an effective, motivating, and meaningful context for reading comprehension instruction (Hwang et al., 2022; Maerten-Rivera et al., 2016).

Each lesson consisted of five instructional phases that were ordered according to the Gradual Release of Responsibility instructional model (e.g., Fisher & Frey, 2021). During explicit instruction, teachers briefly introduced or reviewed one text structure. Then, teachers followed a modeling script and demonstrated how the structure could be recognized (i.e., how to look for text-structural cues), and how reading strategies (e.g., predicting, summarizing) should be applied in that specific structure. In three lessons, videoclips were provided in which a stronger and a weaker model read the same text while thinking aloud. After this, students and teachers engaged in a moment of reflection or abstraction (Schutz & Rainey, 2020), where they discussed their

teachers' reading approach (e.g., What strategies did I apply during reading, and why?) or compared the video clips (e.g., Which student had the best approach, and why?). Such an active reflection after modeling explicitly focuses students' attention at crucial aspects of the reading approach, which seems to be related to improved student outcomes, at least in the area of writing (Braaksma et al., 2002).

Then, the role of the teacher was faded out: after guided practice, students collaborated in groups to read the remainder of the text and work on a structure-based summarization or recall task. This collaborative activity was often designed as a jigsaw activity: students first worked in expert groups on one part of the text, and then formed mixed groups to exchange information with students who read another (part of the) text, to finally complete a joint product (Aronson, 1978). The temporary knowledge disbalance that is created between students in the mixed groups is assumed to create optimal conditions for collaborative learning (e.g., positive interdependence, promotive interaction; Johnson & Johnson, 2017; Kagan, 2021).

Before students read the text, the teachers were prompted to explicate the ultimate goal of the reading activity, such as participating in a quiz or writing a letter. The joint product of the collaborative task was often a structure-based (schematic) summary which then became input for a writing task, a knowledge test, or an oral task (e.g., classroom discussion, short pitch). This created a functional context for reading and summarizing the text, which might positively affect reading comprehension and motivation (Van Ammel et al., 2021; Britt et al., 2018).

Intervention lessons ended with an individual activity, often a writing task (i.e., writing a paragraph in a specific structure) or a metacognitive reflection task (e.g., Can you provide tips for creating a good Venn diagram? Can you think of other situations where you have to read chronological texts?) in order to explicitly teach for transfer (Patton et al., 2022).

The control condition consisted of business-as-usual lessons (BAU). In these lessons, students followed their regular reading comprehension curriculum (e.g., Nieuwsbegrip [Understanding news], Leeslink [Reading link]). These highly-scripted curricular materials emphasize the practice of reading strategies related to different phases of reading (before, during, after reading) by providing texts with questions, but generally lack explicit instruction about informational text structures (Bogaerds-Hazenberg et al., 2022). The observed BAU lessons focused on procedural strategy knowledge, such as how to make predictions before reading a text. The main distinguishing feature of BAU versus TOP lessons was the lack of explicit attention to informational text structures.



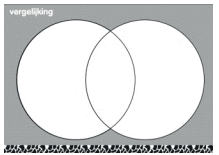
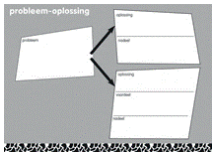
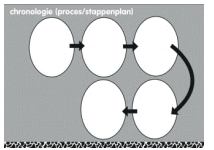
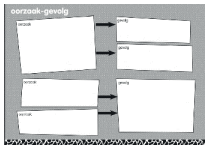
**PEDAGOGISCHE STUDIËN**
https://doi.org/10.59302/ps.v101i1.18783

https://doi.org/10.59302/ps.v10i1i1.18783



***Table 2***

Intervention program TOP

| Introduction<br>Lesson 1 | Compare-contrast<br>Lesson 2 & 3 | Problem-solution<br>Lesson 4 & 5 | Chronology<br>Lesson 6, 7 & 8 | Cause-effect<br>Lesson 9 & 10 |
|---|---|---|---|---|
| Texts<br>Organ systems<br>Seeds and spores<br>Jaguars or cheetahs? | Texts<br>Mustelids, rodents, or pack animals.<br>Viruses or bacteria?<br>North or South Pole?<br>The Olympics: now and then. | Texts<br>Historic inventions<br>Plastic soup | Texts<br>Recipe<br>Sewerage<br>Cocoa to chocolate<br>Canals in Utrecht | Texts<br>The Himalayas<br>Impact of El Niño |
| Activities<br>Sorting tasks | Activities<br>Write summary<br>Prepare a pitch | Activities<br>Write summary<br>Prepare a pitch<br>Write problem-solution text | Activities<br>Summary and quiz<br>Write persuasive letter<br>Write chronological text | Activities<br>Write summary<br>Play a class game<br>Information poster |
| |  |  |  |  |

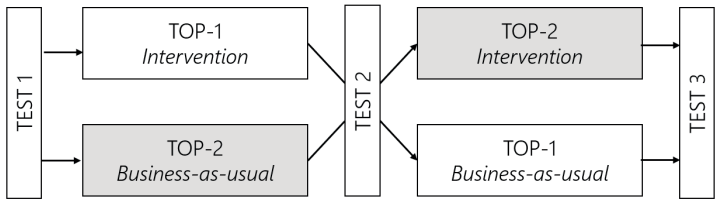
## 2.2 Sample and research design

Ten upper-elementary school teachers (80% female) in seven Dutch primary schools volunteered to participate in the current study. They had on average 9.3 years of teaching experience (*SD* = 5.9) and taught in grade 4, grade 5, or in a mixed grade (5/6). Most schools were religiously affiliated (3 Catholic, 2 Protestant, 2 public schools), which is representative of the Dutch school system. The public schools participated with respectively two and three classes; the other schools with one class. Overall, 203 students (55% female) participated in our study: 114 fourth graders (divided over four classes), 91 fifth graders, and 22 sixth graders (divided over three fifth grade classes and three mixed grade classes). The average class size was 20.7 (*SD* = 6.1).

Classes and teachers were randomly assigned to one of two groups, as we used a switching-panels replication design (Shadish et al., 2002; Figure 1) with three measurement occasions (T1; T2; T3). The TOP-1 group implemented the intervention (TOP) between T1 and T2 and followed business-as-usual lessons (BAU) between T2 and T3. The TOP-2 group started with their regular reading curriculum, and implemented the intervention program between T2 and T3.

***Figure 1***

Switching panels replication design with two groups: TOP-1 and TOP-2






**PEDAGOGISCHE STUDIËN**

https://doi.org/10.59302/ps.v101i1.18783

Table 3 provides various class-level and student-level features per group and grade. Four classes somewhat overrepresented students with a lower socio-economic and migration background; these were equally distributed over both groups. The groups were highly comparable; there were no significant differences across groups with regards to gender ($F(3, 197) = 1.90$, $p = .13$), or the average class size ($F(3, 6) = 2.17$, $p = .19$). Students with dyslexia ($n = 14$) were equally distributed over the intervention groups and grades ($F(3, 197) = 1.35$, $p = .26$). One exception was the fact that the teachers of the fourth graders in the TOP-2 group on average had much less teaching experience than those of the fourth graders in the TOP-1 group ($\Delta M = 10.67$ years, se = 3.77, $p = .03$).

*Table 3*

Characteristics per group and grade

| | TOP-1 | | TOP-2 | |
|---|---|---|---|---|
| | Grade 4 | Grade 5+ ** | Grade 4 | Grade 5+ |
| Students | 48 | 60 | 40 | 53 |
| Classes | 2 | 3 | 2 | 3 |
| Class size | 21.0 (0.0) | 21.0 (9.6) | 20.0 (4.2) | 21 (3.6) |
| Male students (%) | 54 | 47 | 43 | 57 |
| Teacher experience (years) | 16.0 (5.7) | 11.0 (6.1) | 3.0 (2.8) * | 7.3 (2.1) |

Note. TOP-1: text structure lesson series before business-as-usual; TOP-2: vice versa.

* Means differed significantly between groups, with $p < .05$.

** Grade 5+ concerns 91 fifth graders and 22 sixth graders who participated due to mixed grade groupings.

### 2.3 Measures

Students completed various tests at three measurement occasions. The tests were administered one week prior to the start of the experiment (T1), and within two weeks after the end of the intervention (T2 and T3). Teachers spread the tests over three sessions within one week. Students could work without a time limit, but even so, some students failed to complete one or two tests at the first ($n$ = 12), second ($n$ = 10), or third ($n$ = 13) measurement occasion.

*Text structure test*

The text structure tests that we designed mainly consisted of multiple-choice questions ($n$ = 10 for T1, $n$ = 9 for T2 and T3) that required students to recognize the underlying structure of text fragments or to make structure-based predictions based on a title, or to complete a main idea sentence. In addition, for two content-area texts ($M_{length}$ = 207.5 words, $SD$ = 33.0), students received a list of numbered main ideas and a graphic organizer in which they had to put the numbers of the main ideas in the right boxes. This list of numbered main ideas always contained two incorrect main ideas as distractor items; students were instructed that two statements would be left after completing the graphic organizer. The graphic organizer questions were inspired by sorting tasks that were used in previous studies to measure situation model comprehension (Land, 2009). Students completed a problem-solution and a cause-effect chart at T1, a timeline and Venn diagram at T2, and a timeline and problem-solution chart at T3. Students received 1 point for each correct box. The maximum number of points to be earned per graphic organizer varied from 6 to 10.

For scaling purposes, the text structure tests were piloted among fourth and fifth graders in three non-participating schools ($n$ = 98). The analyses showed that the test items had similar mean $p$-values, but that *rit*-values were rather weak. Therefore, some malfunctioning items were deleted (two items on T1, two on T2, and five on T3). The final reading tests were comparable in difficulty; $p$-values (.59, .59, and .58) and variances (.24 for all tests), but still not completely parallel in terms of their *rit*-values (.55, .30, and .59). Therefore, we concluded that the text structure tests could be used to reliably detect interaction effects resulting from the experimental manipulation, but that changes in scores between measurement occasions should not be interpreted as absolute growth in student abilities. Reliability analysis produced acceptable to good alpha-coefficients ($\alpha$ = .73, $\alpha$ = .80, and $\alpha$ = .86).

*Standardized reading test*

The standardized reading tests contained expository passages and one narrative passage with multiple choice items and open-ended questions that pertained to literal text base comprehension or situation-model representation (Table 4). The test items ($n$ = 20 per occasion) were assembled from a database of yet unpublished PPON-items that are used for Dutch periodical nationwide reading assessments (see Kuhlemeier et al., 2014). All of these items had already been piloted on a large scale with third and sixth graders. As we wanted to interpret the changes in test scores between measurement occasions as actual growth in reading skills, we created three parallel reading tests by assembling items so that they were comparable in average $p$-values (.68, .68, and .65), variances (.22, .22, and .23), and *rit*-values (.33 at all three occasions). Reliability analyses produced more or less acceptable alpha-coefficients per measurement occasion ($\alpha$ =.59, $\alpha$ =.65, and $\alpha$ =.64). We had to delete one item at T2, as analyses showed this item did not function properly.



PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/ps.v101i1.18783

***Table 4***

Overview of the standardized reading test at three measurement occasions

| T1 | T2 | T3 |
|---|---|---|
| Informational text (description) | Informational texts (1: description, 2: unclear structure) | Informational text (problem-solution) |
| Informational text (problem-solution) | Instruction text (chronological structure) | Informational text (± problem-solution) |
| Informational text (±compare-contrast and ±chronological structure) | Informational text (±chronological structure) | Informational text (unclear structure) |
| Narrative text | Narrative text | Narrative text / Interview |

Note. The indication ± means that only a few paragraphs in the text displayed to some extent a certain text structure.

*Summary*

At each measurement occasion, students summarized an informational text, consisting of an introduction to a problem, and three paragraphs with possible solutions ($M_{length}$ = 302.2 words, SD = 12.3). The text topics related to geography (traffic jams) on T1, history (the Dutch water defense system) on T2, and biology (venomous animals) on T3. Dotted lines on the answer sheet provided implicit guidance about the desired summary length.

A scoring protocol was developed for each text by the first and second author. Summaries were first parsed into idea units (i.e., sentences that contained at least one verb), and each idea unit was then classified as main idea (1 point) or detail (0 points). For each text, students could score a predefined number of main ideas that was comparable across measurement occasions, and related to the crucial parts of a problem-solution text, such as Problem, Solution-1, Solution-2, and so on. The percentage of correctly included main ideas was used as outcome measure. Students did not receive points for redundant or false information. In order to check the reliability of this coding procedure, for each summary task, 32 summaries (16%) were also scored by one of two trained research-assistants, which revealed substantial agreement between raters (87%, 90%, and 81%). Several differences between the raters were discussed in the research team. For one summary task (T3), the scoring protocol was slightly refined, and the first author corrected the rating of T3 accordingly. Reliability analyses produced more or less acceptable alpha-coefficients ($\alpha$ =.58, $\alpha$ =.79, and $\alpha$ =.70).



PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/ps.v101i1.18783

*Writing*

At each measurement occasion, students completed one writing task, taken from the Tekster intervention program (Bouwer & Koster, 2016). The three writing tasks were comparable with regard to the intended audience and the communicative goal: at each measurement occasion, students had to write a persuasive letter (e.g., Convince your teacher that you and your classmates need a classroom pet). Two raters independently assessed the text quality on a continuous rating scale (0-5) with five benchmark essays that represented the range of writing quality that can be expected of fourth and fifth graders (Bouwer & Koster, 2016). Interrater reliability analysis showed agreement percentages of 80%, 86%, and 85%. Text quality scores were obtained by calculating the mean score of the two raters.

### 2.4 Treatment fidelity

As it is important to gain insight into teachers' enactment of reading interventions (e.g., Beerwinkle et al., 2018; Bohaty et al., 2015; Okkinga et al., 2021), we collected some data related to treatment fidelity. We observed two

intervention lessons per teacher ($n$ = 20) and one business-as-usual lesson per teacher ($n$ = 10) to carefully describe the content and instructional approach of the curriculum. For the control condition, it served as an extra check that there was no explicit text structure instruction. An observation sheet was used to evaluate for each lesson phase (i.e., explicit instruction, modeling and reflection, guided practice, collaborative learning, individual practice, closure) its occurrence and duration (in minutes), and to check the quality of key components (1: implemented as intended; 0: not implemented as intended). For example, for collaborative learning, it was checked whether the teacher first created expert and then mixed groups (which is required for the jigsaw tasks in the intervention program). Lessons could not be videotaped, but the validity of the observation scheme was increased by formulating concrete events that were evaluated on occurrence and duration, and by leaving room for a general impression and for remarks, so that difficult decisions could be discussed afterwards (Miles et al., 2013). In addition, all teachers completed a questionnaire on their regular reading curriculum.

Social validity data were gathered through teacher entries in logbooks in which they evaluated each lesson, and rated on a Likert scale their lesson appreciation, and the estimated level of difficulty of the lesson. At the end of the TOP program, we conducted retrospective interviews with the participating teachers. Due to unforeseen circumstances, we were only able to collect logbook ($n$ = 6) and interview data ($n$ = 6) of half of the teachers.



https://doi.org/10.59302/ps.v101i1.18783

### 2.5 Analysis

Multilevel models were applied to the data. For the analysis of effects on the summarization task and the writing task, linear multilevel models were applied on sum scores (i.e., the total score on each test per occasion) within a hierarchical three-level structure, with measurement occasions (T1, T2, T3; level 1) nested within students (level 2), who were nested within classes (level 3) (Hox et al., 2017). By stepwise model fitting (see Appendix), we added as fixed effects Time (three measurement occasions: T1, T2, T3), Group (TOP-1 or TOP-2), Grade (4 or 5+) and their interactions to estimate means per group and measurement occasion, with variation within and between students and classes as random effects. We examined the immediate effects of the intervention (T1-T2 for TOP-1; T2-T3 for TOP-2), and the delayed effects of the intervention as well (T2-T3 for TOP-1). In the case of interaction effects, custom hypotheses were applied to localize the effect (between T1-T2, and/or between T2-T3).

Generalized binominal multilevel models were applied to the binominal data of the text structure test and the standardized reading test. Because of the heterogeneous nature of the items, an additional level of items nested within measurement occasions was used, resulting in a hierarchical four-level

structure (i.e., score on each single item per task per student per measurement occasion), with items (level 1) nested within measurement occasions (T1, T2, T3; level 2), nested within students (level 3), who were nested within classes (level 4). Following a similar approach to estimate means per group, we examined the immediate effects of the intervention (T1-T2 for TOP-1; T2-T3 for TOP-2), as well as the delayed effects of the intervention (T2-T3 for TOP-1). Variances for items and students within classes, and the variance between classes were modeled as random effects.

## 3 Results

### 3.1 Treatment fidelity

On a 10-point scale, teachers gave a positive overall rating of the intervention program ($M$ = 8.14, $SD$ = 1.03). In particular, they appreciated the modeling scripts and the structure-specific graphic organizers. Teachers followed the intervention program quite strictly, although some issues related to conditional knowledge were observed: in less than half of the observed intervention lessons (45%), teachers explained when a specific reading approach would be most useful, and one third of the teachers (30%) did not initiate the intended metacognitive reflection after modeling. Teachers felt able to provide instruction ($M$ = 3.19, $SD$ = 1.26; on a five-point scale), but rated the lessons as somewhat difficult for their students ($M$ = 3.73, $SD$ = 0.66).

The TOP lessons ($M$ = 55.2 minutes, $SD$ = 9.9) were longer than regular reading lessons ($M$ = 47.1, $SD$ = 9.0). The average duration of explicit instruction and guided practice was quite similar, but during TOP lessons, more time was allocated to modeling ($M$ = 8.83, $SD$ = 3.05) and collaborative learning ($M$ = 12.01, $SD$ = 0.93) than during business-as-usual ($M$ = 1.60, $SD$ = 1.99, and $M$ = 3.00, $SD$ = 4.52 respectively). Business-as-usual instruction focused on activating prior knowledge and, to a lesser extent, on training reading strategies.

Students in BAU lessons spent more time on individual practice (42% of total lesson time; $M$ = 19.80, $SD$ = 14.12) than during TOP lessons (16% of total lesson time; $M$ = 9.00, $SD$ = 6.75). In half of the observed TOP lessons, teachers did not create expert and mixed groups for the collaborative learning task – a structure that promotes positive interdependence, individual accountability, and a high level of interactions among students – but instead, they turned the jigsaw tasks into simple pairwise activities in which these key principles of effective collaborative learning could no longer be guaranteed. Observations and logbooks also revealed that teachers often skipped the individual activities and feedback at the end of intervention lessons, which may be linked to teachers' comments about the relatively long lesson duration. Compared to the TOP-1


PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/ps.v101i1.18783

teachers who implemented the program first, the TOP-2 teachers appeared to devote 35% more time to explicit instruction ($\Delta M$ = 2.22, se = 1.52, $p$ = .04) and 9% more time to individual practice ($\Delta M$ = 0.78, se = 3.13, $p$ = .03).

### 3.2 Immediate and delayed effects

The following sections discuss the immediate and delayed intervention effects in both groups per outcome measure. As the correlations between the different outcome measures appeared relatively weak (all $r \leq .53$), we discuss the effects for each outcome measure separately (see Appendix, Table A1 and A2 for more details). The model fitting procedures, parameter estimates, and the estimated means can be found in the Appendix as well (Tables B-E).

*Text structure test*

Figure 2 provides an overview of the estimated means per group and grade on the text structure test at three measurement occasions. There were main effects of Grade ($F$(1, 10556) = 7.47, $p$ = .006; which can be interpreted as known group validity) and Time ($F$(2, 10556) = 3.40, $p$ = .03), and interaction effects of Time and Group ($F$(2, 10556) = 21.34, $p$ < .001), Grade and Time ($F$(2, 10556) = 3.11, $p$ =.04), and a three-way interaction of Grade, Group and Time ($F$(2, 10556) = 4.62, $p$ = .01).

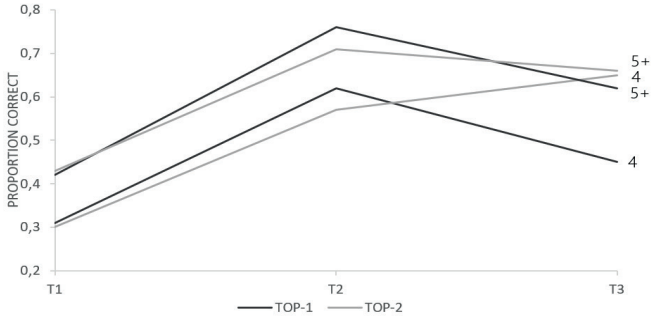
For the TOP-1 group, there was no demonstrable immediate and therefore no delayed effect of the intervention on the text structure test; at T2, the groups did not differ significantly (all $p$s > .05). For the TOP-2 group, there was an immediate effect of the intervention, with an effect size of $d$ = 0.50, but only for fourth graders. Different from the other groups, they scored relatively higher at T3 with respect to T2, so that they outperformed fourth graders in the TOP-1 group ($\Delta M$ = .20, se = .07, $p$ = .006), and scored comparable to older students at T3 ($\Delta M$ = .01, se = .07, $p$ = .87).



PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/ps.v101i1.18783

***Figure 2***

Text structure test: Estimated means per grade and group



*Standardized reading test*

Figure 3 shows the estimated means for the standardized reading test at three measurement occasions (T1, T2, T3). There was a main effect of Grade ($F$(1, 7418) = 4.30, $p$ = .04; known group validity), as well as interaction effects of Time and Group ($F$(2, 7418) = 4.53, $p$ = .01), Grade and Time ($F$(2, 7418) = 9.81, $p$ < .001), and a three-way interaction between Grade, Time, and Group ($F$(2, 7418) = 3.34, $p$ = .04).

For the TOP-1 group, no demonstrable immediate effect and as a consequence no delayed effect could be shown on the standardized test: neither of the groups improved between T1 and T2, so that both groups did not differ at T2 (all $p$s > .05).

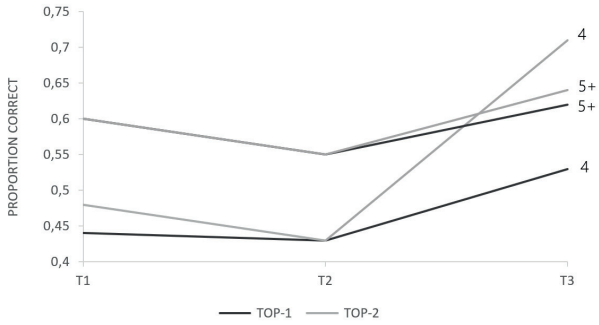
For the TOP-2 group, we found an immediate effect of the intervention, but only for fourth graders, with an effect size of $d$ = 0.53. Between T2 and T3 they significantly improved ($\Delta M$ = .28, se = .17, $p$ = .045), so that at T3, they outperformed fourth graders in the TOP-1 group ($\Delta M$ = .18, se = .07, $p$ = .01), and scored comparable to the fifth and sixth graders at T3 ($p$ = .98). Overall, the variance between items is relatively high on the standardized reading test ($S^2$ = 2.89, se = .80), suggesting that items might have been too heterogeneous to demonstrate effects.

***Figure 3***

Standardized reading test: Estimated means per grade and group



https://doi.org/10.59302/ps.v101i1.18783




*Summarization*

Figure 4 shows the estimated means for the summarization task at three measurement occasions. For summarization – and specifically the main ideas included in students' summaries – we found a main effect of Grade ($F$(1, 20) = 14.12, $p$ = .001); fourth graders included 14% fewer main ideas in their summaries than older students ($\beta$ = -.14, se = 3.91; known group validity). In addition, a main effect of Time was found ($F$(2, 553) = 13.90, $p$. <.001), and an interaction effect of

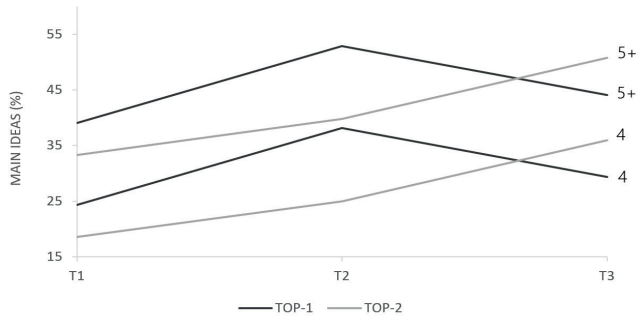
Time and Group ($F(2, 553) = 9.38$, $p < .001$).

For the TOP-1 group, there was no immediate intervention effect, and as a consequence, no delayed effect. Even though TOP-1 students scored better on T2 with respect to T1 ($\Delta M = 13.82$, se = 3.07, $p < .001$), the interaction effect of Time and Group between T1 and T2 was not significant ($\Delta M = -7.61$, se = 4.65, $p = .10$).

For the TOP-2 group, there was an immediate effect of the intervention, with an effect size of $d = 0.48$. TOP-2 students scored better on T3 with respect to T2 ($\Delta M = 11.68$, se = 5.10, $p = .005$), while the TOP-1 group scored lower ($\Delta M = -8.82$, se = 3.15, $p = .005$). The interaction effect of Time and Group between T2 and T3 was significant ($\Delta M = 19.67$, se = 4.64, $p < .001$).

***Figure 4***

Summarization: Estimated means per grade and group (% main ideas)




*Writing*

Figure 5 shows the estimated means for writing at three measurement occasions. A main effect of Grade could be demonstrated ($F(1, 9.4) = 31.48$, $p = .003$), showing that fourth graders scored lower than students in the other grades on writing ($\beta = -0.42$, se = 0.11, $p < .001$; known group validity). In addition, an interaction effect of Time and Group ($F(2, 561) = 4.31$, $p = .01$) was found.

For the TOP-1 group, there was an immediate effect of the intervention, with an effect size of $d = 0.33$. That is, students in this group scored higher on T2 than on T1 ($\Delta M = 0.30$, se = 0.12, $p = .02$), while the TOP-2 group scored lower ($\Delta M = -0.22$, se = 0.14, $p = .11$). As a result, TOP-1 students outperformed students in the regular curriculum at T2 ($\Delta M = 0.42$, se = 0.14, $p = .004$). This interaction effect of Time and Group between T1 and T2 was significant ($\Delta M = 0.53$, se = 0.19, $p = .005$).

For the TOP-2 group, we also found an immediate effect of $d = 0.39$. The



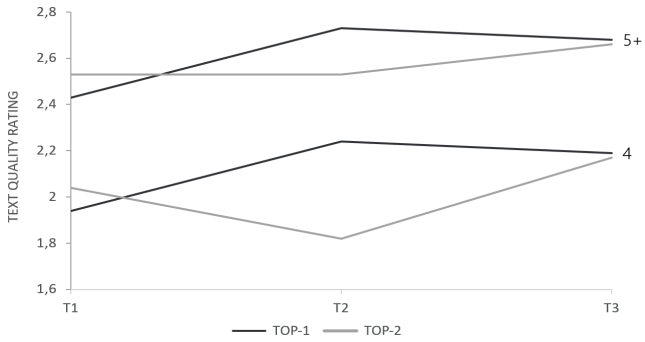


PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/ps.v101i1.18783

TOP-2 group scored higher on T3 with respect to T2 ($\Delta M$ = 0.36, se = 0.14, $p$ = .009), while students in the TOP-1 group remained quite stable in their writing between T2 and T3 ($p$ > .05). This interaction effect of Time and Group between T2 and T3 was significant ($\Delta M$ = 0.40, se = 0.18, $p$ = .03).

***Figure 5***

Writing: Estimated means per grade and group (rating 0-5)




## 4 Conclusion and Discussion



PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/ps.v101i1.18783

This quasi-experimental study evaluated the effectiveness of a text structure intervention (TOP) that was developed for Dutch students in the upper elementary grades. With a switching-panels replication design, we evaluated the intervention effects in two iterations (TOP-1 and TOP-2) on reading comprehension, summarization, and writing. For most outcome measures, one out of four possible immediate intervention effects could be demonstrated. Only fourth graders in the TOP-2 group outperformed other students on all outcome measures, directly after finishing the intervention program: the text structure test ($d$ = 0.50), the standardized reading test ($d$ = 0.53), and the summarization task ($d$ = 0.48). On the writing task, effects were demonstrable across grades in both iterations of the intervention ($d$ = 0.33 and $d$ = 0.39).

The effects that were demonstrated for the fourth graders in the TOP-2 group resonate with meta-analytic research: text structure instruction can have a positive effect on text structure knowledge, summarization, and students' performance on reading comprehension tests (Bogaerds-Hazenberg et al., 2021; Hebert et al., 2016; Pyle et al., 2017). Knowledge about text structure provides students with a tool to identify main ideas and organize these in a meaningful framework, which is crucial for text comprehension and for getting the gist of the text (e.g., Meyer et al., 1980; Stevens & Vaughn, 2021; Strong, 2020). For this reason, researchers have argued that text structure instruction should be a crucial component of reading and summarization instruction (Stevens & Vaughn,

2021; Wijekumar et al., 2019), in particular as current textbooks for reading often display a poor coverage of text structures and summarization skills (Beerwinkle et al., 2018; Bogaerds-Hazenberg et al., 2022; Wijekumar et al., 2019).

The effect on writing in both iterations of the intervention, might suggest that students benefited from the text structure instruction and the writing tasks. This lends support for the assumption that knowledge of text structures provides readers with a useful frame to organize information in their mind, and writers with a means to organize information on paper (Dickson, 1999). In the current study, the type of writing task (a persuasive letter) was not closely aligned with the intervention content, as we hoped to measure students' skills in informational writing through the summarization task. Yet, in a future study, it might be worthwhile to examine the effect sizes on a writing task that is more closely aligned with the intervention content, and would prompt students to write an informational text with a specific structure (see Hebert et al., 2018; Strong, 2020).

In the current study, most findings are not generalizable, as the effects were often only demonstrated for one particular group of students: the fourth graders in the second intervention group. This remarkable outcome might not have been discovered with a single quasi-experimental design, which illustrates the added value of the switching-panels replication design: it allows for studying the replicability of effects (Shadish et al. 2002), even though it might generate more questions than answers. Therefore, we will discuss our findings in the light of a critical inspection of test alignment and quality, intervention content, and fidelity of implementation.



PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/ps.v101i1.18783

### 4.1 Test alignment and quality

First, we need to reflect on the degree of alignment of the outcome measures with the intervention content. It is remarkable that not all intervention groups performed better at the text structure test right after finishing the intervention program. The text structure test may have turned out less aligned with intervention content than intended. Several test items did not directly measure students' text structure knowledge, but required multiple steps of reasoning. For example, instead of asking students which cue words can be found in cause-effect texts, or instead of having them fill in a cloze task, students were asked: 'Peter reads a text on the consequences of unemployment; what signaling words will he probably encounter in the text?' This requires students to infer in a first step that this text probably has a cause-effect structure, and in a second step, recall which signaling words are characteristic of cause-effect structures. Hence, the text structure test turned out to be a mid-transfer test at best (Clemens & Fuchs, 2021). In addition, several graphic organizer items were inspired by a study among secondary school students (Land, 2009), but these items did not perfectly resemble the graphic organizers of the intervention program. Studies

with better alignment in this respect often get clearer results (e.g., Strong, 2020). Moreover, the text structure test might have measured more than one construct: it contained graphic organizer items as well as items that targeted explicit knowledge of text structure characteristics. The correlations between these types of items was relatively low ($r$ = .15 at T1, $r$ = .25 at T2, and $r$ = .28 at T3). Alternative tests can be found in international studies. For instance, recent research has revealed that a text structure identification test in which students identify the structure of paragraph-length texts is a valuable approach to measure one construct: text structure awareness (TSIT; Strong, 2023). The development of a Dutch equivalent might be worthwhile. In addition, one could consider using a more indirect approach of measuring text structure knowledge, for example by using cloze tasks on signaling word knowledge, and (multiple-choice) questions for main idea identification (Wijekumar et al., 2017).

The lack of consistent outcomes on the standardized reading test across intervention groups also raises questions with regard to alignment. It probably reflects the fact that standardized reading tests measure more strategies and skills than those directly related to insight into text structures (Hebert et al., 2016; Leslie & Caldwell, 2009; van den Broek & Kremer, 2000). Our standardized reading tests also contained literal comprehension questions and questions pertaining to narrative texts, which might have made the test less sensitive to readers' situation model reading skills for informational texts; the skills that are assumed to be primarily affected by text structure instruction (Kintsch, 2004; Meyer & Ray, 2011; Wijekumar et al., 2019; Williams, 2018). Therefore, the standardized test should be considered as a far transfer test (Clemens & Fuchs, 2021).

Another issue that warrants a critical reflection is the quality of the tests in relation to the research design. In order to examine growth patterns and effect maintenance, a switching-panels replication design requires three perfectly parallel constructed tests for each measurement occasion (i.e., tests with equal means, observed and true score variances). If this condition is met, changes in scores between measurement occasions can unequivocally be interpreted as growth. As there were no Dutch parallel reading tests freely accessible, all reading tests had to be developed from scratch, and all tests needed to be perfectly parallel. Despite our attempts to create three parallel reading tests from a large item data base (Kuhlemeier et al., 2014), and to scale and pilot the researcher-designed text structure test in three non-participating schools, the tests did not turn out as perfectly parallel, and substantial between-item and error variance was detected. As a result, it was impossible to analyze growth patterns, and we had to base our data analysis on interaction effects. The power of a statistical analysis based on interaction effects is lower than one based on main effects of growth. In the light of a limited test quality, it might have been too ambitious to prove intervention effects in two iterations, over and above business-as-usual gains.


https://doi.org/10.59302/ps.v101i1.18783

### 4.2 Intervention content

Second, the limited effectiveness of the text structure intervention also warrants a critical reflection with regard to the intervention duration and content. Students had to learn about four text structures in ten one-hour lessons divided over six weeks. Text structure interventions with more favorable results in the US often had a higher intensity and/or intervention duration, such as fifty lessons during a whole schoolyear (Williams et al., 2016), four lessons a week during eight weeks (Strong, 2020), or one web-based lesson each week during seven months (Wijekumar et al., 2017). In another study among Dutch primary school students, the effects of a reading intervention only showed up after two years of teaching (Droop et al., 2016). On the other hand, our intervention of ten hours was only marginally shorter than what Pyle and colleagues (2017) identified as the most favorable duration of text structure interventions (i.e., 11-20 hours). It is more likely that intervention duration in itself does not fully explain the limited results, but that the complexity of the intervention in relation to its duration should be considered (Hebert et al., 2016).

As the intervention duration was relatively short, there might have been relatively few opportunities for scaffolding and repetition. For example, the intervention by Strong (2020) provided ample opportunities for students to gradually learn how to construct and complete text structure graphic organizers by themselves, as an active construction of graphic organizers is related to increased intervention effectiveness (Bogaerds-Hazenberg et al., 2021). As we did not want to overload the lesson program of TOP, we might not have included a sufficient number of such tasks for students to practice on their own, and instead mainly offered them in the context of collaborative learning. This might have limited the effectiveness, as for a full gradual release of responsibility, students should also practice reading skills without direct support of their peers (Fisher & Frey, 2021).

For some outcome measures, we found a positive intervention effect among fourth graders in the TOP-2 group, but not among fifth graders. One might hypothesize that this is related to a higher knowledge level on behalf of the fifth graders: possibly, they already learned more about text structures in the regular curriculum. However, this explanation is unlikely, since in both grades, teachers and teaching materials pay very limited attention to text structure (Bogaerds-Hazenberg et al., 2022). Hence, for all students, the intervention differed from their regular reading curriculum in terms of content (e.g., learning about four informational text structures, structure-specific strategy use), and instructional approach (e.g., collaborative learning activities).



PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/ps.v101i1.18783

### 4.3 Implementation fidelity

We cannot pinpoint the exact reason for the fact that the fourth graders in the TOP-2 group were the only ones to demonstrate all of the hypothesized effects. It is not likely that the lack of effects in the other groups is due to social validity issues: all teachers were very positive about the intervention program and the opportunities it provided to integrate reading comprehension instruction with content-area subjects (see also Hwang et al., 2021; Maerten-Rivera et al., 2016). Zooming in on the teachers of the group who did demonstrate all the effects, it was found that they had on average less teaching experience, and spent slightly more time on explicit instruction and individual practice than the other teachers. Generally speaking, young and less-experienced teachers rely more on teaching materials (Valencia et al., 2006), but this could be a good thing in case high-quality curricular materials are used. However, we should be very cautious in drawing strong conclusions, as the number of participating teachers is too limited.

Overall, teachers' implementation of the TOP lessons was not completely as intended, which was observed across teachers in all groups. This is not surprising, as many elements of the TOP program differed from the regular reading curriculum (Bogaerds-Hazenberg et al., 2022), and one single training session may not have sufficed. Most issues were related to implementing a GRR: the collaborative learning task was often transformed into simple pairwise work, which undermined the principles of structured collaborative learning. In fact, teachers' implementation undermined the knowledge disbalance that is purposefully created in jigsaw tasks as a means to promote interactions and positive interdependence (Johnson & Johnson, 2017; Topping et al., 2017). That is, instead of having different student experts in mixed groups who exchange information, the students who worked together in simple pairs had read all the information, so that there was no need to really discuss the text content with their peers in order to complete the assignment.

In addition, individual activities were regularly skipped. This might have been be affected by time management and scheduling issues (Gillies & Boyle, 2010; Hebert et al., 2018), as well as by limited teacher knowledge and skills related to collaborative learning (Okkinga et al., 2021; Veldman et al., 2020). Although it is tempting to suggest that the lack of a GRR has tempered the effectiveness of TOP, the implementation data are too limited. Therefore, the only solid conclusion at this point is that we need to gain more insight into implementation fidelity by examining the challenges related to implementing a GRR and/or text structure instruction.

All in all, the promising finding in one of the intervention groups suggests that text structure instruction could foster reading-to-learn skills in the upper elementary grades, but various challenges have to be overcome in the Dutch


https://doi.org/10.59302/ps.v101i1.18783

context. We recommend future research on implementation issues and/or teacher knowledge in order to gain more insight into the support that teachers need for an effective implementation of text structure instruction (Beerwinkle et al., 2018; Hebert et al., 2018), as well as research on the development of high-quality instruments for assessing higher-order reading comprehension.

## References


Aronson, E. (1978). *The jigsaw classroom*. Sage Publishing.

Beerwinkle, A.L., Owens, J., & Hudson, A. (2021). An analysis of comprehension strategies and skills covered within Grade 3-5 reading textbooks in the United States. *Technology, Knowledge and Learning, 26*(1), 311-338.

Bogaerds-Hazenberg, S.T.M., Evers-Vermeul, J., & van den Bergh, H.H. (2022). What textbooks offer and what teachers teach: An analysis of the Dutch reading comprehension curriculum. *Reading and Writing, 35*(7), 1497-1523.

Bogaerds-Hazenberg, S.T.M., Evers-Vermeul, J., & van den Bergh, H.H. (2019). Teachers and researchers as co-designers? A design-based research on reading comprehension instruction in primary education. *Journal of Educational Design Research*, *3*(1), 1-23.

Bogaerds-Hazenberg, S.T.M., Evers-Vermeul, J., & van den Bergh, H.H. (2021). A meta-analysis on the effects of text structure instruction on reading comprehension in the upper elementary grades. *Reading Research Quarterly, 56*(3), 435-462.

Bohaty, J.J., Hebert, M.A., Nelson, J.R., & Brown, J.A. (2015). Methodological status and trends in expository text structure instruction efficacy research. *Reading Horizons, 54*(2), 35-65.

Braaksma, M., Rijlaarsdam, G., & van den Bergh, H.H. (2002). Observational learning and the effects of model-observer similarity. *Journal of Educational Psychology, 94*(2), 405-415.

Coté, N., Goldman, S.R., & Saul, E.U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes, 25*(1), 1-53.

Dickson, S. (1999). Integrating reading and writing to teach compare-contrast text structure: A research-based methodology. *Reading and Writing Quarterly, 15*(1), 49-79.

Dutch Inspectorate of Education (2022). *Peil.leesvaardigheid: Einde (speciaal) basisonderwijs* https://www.onderwijsinspectie.nl/onderwerpen/peil-onderwijs/documenten/themarapporten/2022/12/13/peil-leesvaardigheid-einde-basisonderwijs-en-speciaal-basisonderwijs-2020-2021

Fisher, D., & Frey, N. (2021). *Better learning through structured teaching: A framework for the gradual release of responsibility* (3rd ed.). ASCD.

Gillies, R.M., & Boyle, M. (2011). Teachers' reflections of cooperative learning: A two-year follow-up. *Teaching Education, 22*(1), 63-78.

Gubbels, J., van Langen, A.M.L., Maassen, N.A.M., & Meelissen, M.R.M. (2019). *Resultaten PISA-2018 in vogelvlucht*. Universiteit Twente.

Harlaar, N., Dale, P.S., & Plomin, R. (2007). From learning to read to reading to learn: Sub-





https://doi.org/10.59302/ps.v101i1.18783

stantial and stable genetic influence. *Child Development, 78*(1), 116-131.
Hebert, M., Bohaty, J.J., Nelson, J.R., & Brown, J. (2016). The effects of text structure instruction on expository reading comprehension: A meta-analysis. *Journal of Educational Psychology*, *108*(5), 609-629.
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: An Interdisciplinary Journal, 31*(9), 2165-2190.
Helder, A., Van Leijenhorst, L., & van den Broek, P. (2016). Coherence monitoring by good and poor comprehenders in elementary school: Comparing offline and online measures. *Learning and Individual Differences, 48*(1), 17-23.
Hiebert, E.H., & Mesmer, H.A.E. (2013). Upping the ante of text complexity in the Common Core State Standards: Examining its potential impact on young readers. *Educational Researcher, 42*(1), 44-51.
Hwang, H., Cabell, S.Q., & Joyner, R.E. (2020). Effects of integrated literacy and content-area instruction on vocabulary and comprehension in the elementary years: A meta-analysis. *Scientific Studies of Reading, 26*(3), 223-249.
Johnson, D.W., & Johnson, F.P. (2017). *Joining together: Group theory and group skills* (12th Ed.). Pearson.
Kagan, S. (2021). *Pioneering perspectives in cooperative learning*. Routledge.
Kintsch, W. (2004). The construction-integration model of text comprehension and its implications for instruction. In R. Ruddell and N. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 1270-1328). International Reading Association.
Kraal, A., Koornneef, A.W., Saab, N., & van den Broek, P.W. (2018). Processing of expository and narrative texts by low-and high-comprehending children. *Reading and Writing, 31*(9), 2017-2040.
Kuhlemeier, H., Jolink, A., Krämer, I., Hemker, B., Jongen, I., van Berkel, I., & Bechger, T. (2014). *Balans van de leesvaardigheid in het basisonderwijs en speciaal basisonderwijs: Uitkomsten van de peilingen in 2011 en 2012. PPON-reeks 54.* Cito.
Land, J.F.H. (2009). *Zwakke lezers, sterke teksten? Effecten van tekst- en lezerskenmerken op het tekstbegrip en de tekstwaardering van vmbo-leerlingen.* Eburon.
Maerten-Rivera, J., Ahn, S., Lanier, K., Diaz, J., & Lee, O. (2016). Effect of a multiyear intervention on science achievement of all students including English language learners. *The Elementary School Journal, 116*(4), 600-624.
McVee, M.B., Ortlieb, E., Reichenberg, J.S., & Pearson, P.D. (Eds.) (2019). *The gradual release of responsibility in literacy research and practice.* Emerald Publishing Limited.
Meyer, B.J.F., & Ray, M. (2011). Structure strategy interventions: Increasing reading comprehension of expository text. *International Electronic Journal of Elementary Education*, *4*(1), 127-152.
Meyer, B.J.F., Wijekumar, K., & Lei, P. (2018). Comparative signaling generated for expository texts by 4th-8th graders: Variations by text structure strategy instruction, comprehension skill, and signal word. *Reading and Writing: An Interdisciplinary Journal, 31*(9), 1937-1968.



https://doi.org/10.59302/ps.v101i1.18783

Okkinga, M., van Gelderen, A.J., van Schooten, E., van Steensel, R., & Sleegers, P.J.C. (2021). Implementation quality of principles of reciprocal teaching in wholeclassroom settings: a two-year study with low-achieving adolescents. *Reading Psychology, 42*(4), 323-363.

Patton, S.A., Fuchs, D., Hendricks, E.L., Pennell, A.J., Walsh, M E., Fuchs, L.S., Tracy, W.Z., & Haga, L.Y. (2022). An experimental study to strengthen students' comprehension of informational texts: Is teaching for transfer important? *Learning Disabilities Research and Practice, 37*(2), 124-139.

Puzio, K., & Colby, G.T. (2013). Cooperative learning and literacy: A meta-analytic review. *Journal of Research on Educational Effectiveness, 6*(4), 339-360.

Pyle, N., Vasquez, A.C., Lignugaris/Kraft, B., Gillam, S.L., Reutzel, D.R., Olszewski, A., Segura, H., Hartzheim, D., Laing, W., & Pyle, D. (2017). Effects of expository text structure interventions on comprehension: A meta-analysis. *Reading Research Quarterly, 52*(5), 1-33.

Rooijackers, P., van Silfhout, G., Schuurs, U., Mulders, I., & van den Bergh, H. (2020). Lezen en antwoorden bij teksten met vragen: Een cross-sectionele eye-trackstudie onder vwo-leerlingen. *Pedagogische Studiën, 93*(4), 42-58.

Schutz, K.M., & Rainey, E.C. (2020). Making sense of modeling in elementary literacy instruction. *The Reading Teacher*, *73*(4), 443-451.

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

Stevens, E.A., & Vaughn, S. (2021). Using paraphrasing and text structure instruction to support main idea generation. *Teaching Exceptional Children, 53*(4), 300-308.

Strong, J.Z. (2020). Investigating a text structure intervention for reading and writing in grades 4 and 5. *Reading Research Quarterly, 55*(4), 545-551.

Strong, J.Z. (2023). Measuring text structure awareness in upper elementary grades. *Reading and Writing Quarterly, 39*(1), 16-35.

Topping, K., Buchs, C., Duran, D., & Van Keer, H. (2017). *Effective peer learning: From principles to practical implementation*. Routledge.

Valencia, S.W., Place, N.A., Martin, S.D., & Grossman, P.L. (2006). Curriculum materials for elementary reading: Shackles and scaffolds for four beginning teachers. *The Elementary School Journal, 107*(1), 93-120.

Van Ammel, K., Aesaert, K., De Smedt, F., & Van Keer, H. (2021). Skill or will? The respective contribution of motivational and behavioural characteristics to secondary school students' reading comprehension. *Journal of Research in Reading, 44*(3), 574-596.

van den Broek, P., Helder, A., Espin, C., & van der Liende, M. (2021). *Sturen op begrip: Effectief leesonderwijs in Nederland.* Leiden University. www.tweedekamer.nl/sites/default/files/atoms/files/ 20220203_onderzoeksrapport_sturen_op_begrip_effectief_leesonderwijs_in_nederland_in_opdracht_van_commissie_voor_ocw.pdf

Veldman, M.A., Van Kuijk, M.F., Doolaard, S., & Bosker, R.J. (2020). The proof of the pudding is in the eating? Implementation of cooperative learning: differences in teachers' attitudes and beliefs. *Teachers and Teaching, 26*(1), 103-117.

Welie, C., Schoonen, R., & Kuiken, F. (2018). The role text structure inference skill plays for eighth graders' expository text comprehension. *Reading and Writing, 31*(9), 2065-2094.





https://doi.org/10.59302/ps.v101i1.18783

Wijekumar, K.K., Meyer, B.J.F., & Lei, P. (2017). Web-based text structure strategy instruction improves seventh graders' content area reading comprehension. *Journal of Educational Psychology*, *109*(6), 741-760.

Wijekumar, K.K., Beerwinkle, A.L., Harris, K.R., & Graham, S. (2019). Etiology of teacher knowledge and instructional skills for literacy at the upper elementary grades. *Annals of Dyslexia, 69*(1), 5-20.

Williams, J.P. (2018). Text structure instruction: The research is moving forward. *Reading and Writing: An Interdisciplinary Journal*, *31*(9), 1923-1935.

Williams, J.P., Kao, J.C., Pao, L.S., Ordynans, J.G., Atkins, J.G., Cheng, R., & DeBonis, D. (2016). Close analysis of texts with structure (CATS): An intervention to teach reading comprehension to at-risk second graders. *Journal of Educational Psychology, 108*, 1061-1077.

Xu, Z., Wijekumar, K., Wang, Q., Irey, R., & Liang, H. (2021). The effects of web-based text structure strategy instruction on adult Chinese ELLs' reading comprehension and reading strategy use. *Language Teaching Research*, online preprint.

Young, M. I. (2017). *The effects of teaching reading using the Gradual Release of Responsibility Model and 4th grade reading.* Trevecca Nazarene University.



https://doi.org/10.59302/ps.v101i1.18783

## Auteurs

**Suzanne Bogaerds-Hazenberg** is gepromoveerd aan de Universiteit Utrecht. Haar onderzoek richtte zich op expliciet tekststructuuronderwijs in de bovenbouw van de basisschool. Momenteel is zij werkzaam aan het Behavioural Science Institute, Radboud Universiteit.
**Jacqueline Evers-Vermeul** is universitair hoofddocent bij het Departement Talen, Literatuur en Communicatie, Universiteit Utrecht.
**Huub van den Bergh** is hoogleraar Didactiek en toetsing van het taalvaardigheidsonderwijs bij het Departement Talen, Literatuur en Communicatie, Universiteit Utrecht.

*Correspondentieadres*: Suzanne Bogaerds-Hazenberg, Behavioural Science Institute, Radboud Universiteit, Thomas van Aquinostraat 4, 6525 GD Nijmegen. E-mail: suzanne.bogaerds-hazenberg@ru.nl

## Samenvatting

**Tekststructuurinstructie in het basisonderwijs: effecten op lezen, samenvatten en schrijven**

Hoewel tekststructuurkennis een belangrijke rol speelt bij diep tekstbegrip, krijgt tekststructuur slechts beperkt aandacht in het Nederlandse basisschoolcurriculum. In deze effectstudie met een switching-panels replication design volgden 201 leerlingen uit groep 6-8 een tekststructuurinterventie (TOP). Tijdens de interventie leerden leerlingen om vier informatieve tekststructuren te herkennen en om de tekst samen te vatten in graphic organizers. Ook werden schrijftaken aangeboden passend bij de tekststructuur. Leerlingen volgden eerst de tekststructuurlessen en daarna weer hun reguliere leesmethode (de TOP-1-groep) of volgden eerst reguliere leeslessen en daarna de tekststructuurlessen (de TOP-2-groep). Op drie meetmomenten maakten leerlingen een tekststructuurtoets, een leestoets, een samenvattingstaak en een schrijfopdracht. Alleen de leerlingen in groep 6 die in de TOP-2-groep de interventie volgden lieten alle verwachte positieve effecten zien met de volgende effectgroottes: tekststructuurtoets ($d$ = 0.50), tekstbegripsvragen ($d$ = 0.53), samenvatting ($d$ = 0.48). Voor alle leerlingen bleek de interventie een positief effect te hebben op schrijfvaardigheid ($d$= 0.33 en $d$ = 0.39). De uitkomsten worden besproken in het licht van implementatie- en toetskwaliteit.

**Kernwoorden** tekststructuur, begrijpend lezen, primair onderwijs





PEDAGOGISCHE STUDIËN
https://doi.org/10.59302/ps.v101i1.18783

## Appendix

*Correlations*

Table A1 presents the correlations between all outcome measures at M1. As the correlations are relatively weak, the effects on the four outcome measures have been analyzed separately. Table A2 shows the correlations between measurement occasions for all outcome measures, for each intervention group.

***Table A1***

Correlations between all outcome measures (at pretest)

| Test | Text structure test | Summarization | Writing |
|---|---|---|---|
| Standardized reading test | .53 | .44 | .40 |
| Text structure test | | .33 | .27 |
| Summarization | | | .38 |

Note. All correlations are significant ($p < .05$)

***Table A2***

Correlations between scores at T1, T2 and T3. TOP-1 below diagonal and TOP-2 above diagonal

| | Standardized reading test | | | Text structure test | | |
|---|---|---|---|---|---|---|
| | T1 | T2 | T3 | T1 | T2 | T3 |
| T1 | | .59 | .31 | | .47 | .27 |
| T2 | .59 | | .44 | .59 | | .26 |
| T3 | .48 | .36 | | .50 | .49 | |
| | Summarization | | | Writing | | |
| | T1 | T2 | T3 | T1 | T2 | T3 |
| T1 | | .37 | .08 (n.s.) | | .43 | .32 |
| T2 | .41 | | .34 | .59 | | .50 |
| T3 | .28 | .58 | | .58 | .63 | |

*Successive model fitting procedures*

Table B presents the model fitting procedures for summarization and writing. As the text structure knowledge test and standardized reading test were analyzed at item level with generalized linear models, so that the -2LL is no reliable measure for model comparison; instead, for these outcome measures, we used F-ratios of fixed effects to assess model fit (see Table B).

***Table B***

Model fitting for summarization and writing, with random part: classes, students * classes, and error

| | | Comparison | | | |
|---|---|---|---|---|---|
| | -2LL | Models | $\Delta\chi^2$ | $\Delta df$ | p |
| Summarization | | | | | |
| M0 | 5156.26 | - | - | - | - |
| M1: M0 + Grade | 5145.79 | M0 vs M1 | 10.47 | 1 | < .001 |
| M2: M1 + Time | 5120.12 | M1 vs M2 | 25.67 | 2 | < .001 |
| M3: M2 + Group | 5119.19 | M2 vs M3 | 0.93 | 1 | .33 |
| M4: M3 + Time * Group | 5100.74 | M3 vs M4 | 18.45 | 2 | < .001 |
| M5: M4 + Grade * Time | 5097.51 | M4 vs M5 | 3.23 | 2 | .20 |
| M6: M5 + Grade * Group | 5097.24 | M5 vs M6 | 0.27 | 2 | .87 |
| M7: M6+ Grade * Time * Group | 5094.62 | M6 vs M7 | 2.62 | 2 | .27 |
| Writing | | | | | |
| M0 | 1535.71 | - | - | - | - |
| M1: M0 + Grade | 1522.46 | M0 vs M1 | 13.25 | 1 | < .001 |
| M2: M1 + Time | 1517.58 | M1 vs M2 | 4.88 | 2 | .09 |
| M3: M2 + Group | 1515.91 | M2 vs M3 | 1.67 | 1 | .20 |
| M4: M3 + Time * Group | 1507.40 | M3 vs M4 | 8.51 | 2 | .02 |
| M5: M4 + Grade * Time | 1506.88 | M4 vs M5 | 0.52 | 2 | .77 |
| M6: M5 + Grade * Group | 1505.76 | M5 vs M6 | 1.12 | 2 | .57 |
| M7: M6 + Grade * Time * Group | 1504.26 | M6 vs M7 | 1.50 | 2 | .47 |

*Table C*

Fixed effects in the final models per outcome measure

| **Text structure test** | *F* | $df_1$ | $df_2$ | *p* |
|---|---|---|---|---|
| Grade | 7.47 | 1 | 10556 | .006 |
| Time | 3.40 | 2 | 10556 | .03 |
| Group | 0.15 | 1 | 10556 | .70 |
| Time * Group | 21.30 | 2 | 10556 | < .001 |
| Grade * Time | 3.11 | 2 | 10556 | .04 |
| Grade * Group | 0.32 | 1 | 10556 | .57 |
| Grade * Time * Group | 4.62 | 2 | 10556 | .01 |
| **Standardized reading test** | | | | |
| Grade | 4.30 | 1 | 7418 | .04 |
| Time | 0.33 | 2 | 7418 | .72 |
| Group | 0.95 | 1 | 7418 | .33 |
| Time * Group | 4.53 | 2 | 7418 | .01 |
| Grade * Time | 9.81 | 2 | 7418 | < .001 |
| Grade * Group | 0.68 | 1 | 7418 | .41 |
| Grade * Time * Group | 3.34 | 2 | 7418 | .04 |
| **Summarization** | | | | |
| Grade | 14.12 | 1 | 19.8 | .001 |
| Time | 13.90 | 2 | 553.3 | < .001 |
| Group | 0.87 | 1 | 10.5 | .37 |
| Time * Group | 9.38 | 2 | 553.1 | < .001 |
| **Writing** | | | | |
| Grade | 31.48 | 1 | 9.4 | .003 |
| Time | 2.39 | 2 | 561.8 | .09 |
| Group | 1.57 | 1 | 8.4 | .24 |
| Time * Group | 4.31 | 2 | 561.3 | .014 |



https://doi.org/10.59302/ps.v101i1.18783

*Parameter estimates*

Table D shows the estimated means and standard errors for summarization and writing. The dashed line represents the switch in conditions: TOP-1 received the intervention between Time 1 and 2, TOP-2 between Time 2 and 3. Table E presents the estimates and standard deviations for the text structure test and the standardized reading test. As these estimates stem from a generalized linear model, the estimates should be interpreted as logits. For the figures in the running text, these values were transformed in proportions.

*Table D*

Estimated means and standard errors per outcome measure

| Summarization (scale 0-100) | TOP-1 Grade 4 | TOP-2 Grade 4 | TOP-1 Grade 5+ | TOP-2 Grade 5+ |
|---|---|---|---|---|
| Time 1 | 24.33 (4.1) | 18.61 (4.1) | 39.08 (4.1) | 33.36 (4.1) |
| Time 2 | 38.15 (4.0) | 25.00 (4.0) | 52.91 (4.0) | 39.76 (4.0) |
| Time 3 | 29.33 (4.1) | 36.01 (4.0) | 44.09 (4.1) | 50.76 (4.1) |
| $S^2$ between classes | 37.67 (29.9) | | | |
| $S^2$ students in classes | 497.32 (20.6) | | | |
| Writing (scale 0-5) | | | | |
| Time 1 | 1.94 (0.10) | 2.04 (0.11) | 2.43 (0.10) | 2.53 (0.11) |
| Time 2 | 2.24 (0.10) | 1.82 (0.11) | 2.73 (0.10) | 2.31 (0.11) |
| Time 3 | 2.19 (0.10) | 2.17 (0.11) | 2.68 (0.10) | 2.66 (0.11) |
| $S^2$ between classes | 0.81 (0.05) | | | |
| $S^2$ between students | 0.005 (0.009) | | | |

*Table E*

Estimates in logits for the text structure test and for the standardized reading test

| Model result | Text structure test | | Standardized reading test | |
|---|---|---|---|---|
| | Estimate | *se* | Estimate | *se* |
| Grade 4 | 0.66 | 0.44 | 0.90 | 0.54 |
| Grade 5+ | 0.61 | 0.45 | 0.55 | 0.53 |
| Time 1 | -1.46 | 0.56 | -0.97 | 0.68 |
| Time 2 | -0.33 | 0.53 | -1.19 | 0.73 |
| TOP-1 group | -0.81 | 0.30 | -0.77 | 0.29 |
| Time 1 * TOP-1 | 0.88 | 0.19 | 0.58 | 0.20 |
| Time 2 * TOP-2 | 1.02 | 0.17 | 0.79 | 0.22 |
| Time 1 * Grade 5+ | 0.51 | 0.18 | 0.83 | 0.20 |
| Time 2 * Grade 5+ | 0.54 | 0.17 | 0.82 | 0.22 |
| TOP-1 * Grade 5+ | 0.72 | 0.40 | 0.70 | 0.38 |
| TOP-1 * Grade 5+ * Time 1 | -0.74 | 0.24 | -0.51 | 0.27 |
| TOP-1 * Grade 5+ * Time 2 | -0.50 | 0.24 | -0.74 | 0.30 |
| $S^2$ between classes | 0.048 | 0.05 | 0.05 | 0.04 |
| $S^2$ students in classes | 0.66 | 0.08 | 0.40 | 0.06 |
| $S^2$ between items | 2.33 | 0.53 | 2.89 | 0.80 |