Reading and Writing
https://doi.org/10.1007/s11145-025-10721-0

Long-term effects of an SRSD writing intervention for elementary school children

Naymé Salas[1] · **Jofre Ventura**[1] · **Anna Llauradó**[2]

Received: 22 August 2025 / Accepted: 7 October 2025

Abstract

This study investigated the long-term effectiveness of the Self-Regulated Strategy Development (SRSD) program to improve the writing skills of 2nd and 4th grade students. SRSD was used to deliver strategy-based instruction aimed to improve the quality of opinion essays. Although generally effective at posttest, the long-term efficacy of SRSD interventions on written text quality has been researched scarcely in previous large-scale intervention studies. We randomly assigned the schools were 909 Catalan-speaking children (473 boys) attended 2nd (*M* age: 7 years, 5 months) and 4th grade (*M* age: 9 years, 4 months) to a treatment condition or to a business-as-usual control condition. Children in the treatment condition received 11 sessions of SRSD instruction to plan opinion essays. Besides asking children to produce an initial opinion essay to determine baseline (pretest) measures, we collected writing samples at posttest, delayed post-test, mid-term follow-up, and long-term follow-up, which took place at 1 week, 4 weeks, 6 months, and 18 months after instruction, respectively. Results showed that benefits from the SRSD intervention were detectable even after 18 months had passed since the end of the intervention. The magnitude of SRSD benefits decreased over time for all measures, but were higher than pretest scores. Fourth graders retained benefits for longer than second graders for most measures. We have obtained compelling evidence of outstandingly long-lasting effects of SRSD on writing performance. However, our findings also underscore the need for continued support and practice to maintain benefits.

Keywords Writing instruction · SRSD · Maintenance · Longitudinal · Text quality

Extended author information available on the last page of the article

Published online: 28 October 2025

Introduction

Writing is a complex, goal-oriented skill that is one of the main targets of elementary education and beyond. This is because writing has a crucial role in an individual's academic, professional, and personal development. Also, competent writing has been repeatedly shown to enhance learning across the curriculum (Graham et al., 2020). Identifying optimal teaching methods is thus paramount to delivering efficient writing instruction, while it should inform our understanding of the science of writing.

Meta-analytical studies have identified several teaching practices that contribute to improving elementary students' writing skills. In short, most writing instruction methods examined have tended to show a positive impact on writing skills and outcomes. For example, giving feedback to students about their writing, engaging students in prewriting activities, or teaching transcription skills have all been found to be associated with gains in writing competence, including an improvement in overall text quality (Collins et al., 2024; Graham et al., 2012). However, not all teaching practices are equally effective. Graham et al., (2012) found that the most effective writing approach was Self-regulated Strategy Development (SRSD), with an effect size (ES) of 1.17. In contrast, peer or self-feedback, word processing or adopting a process approach yielded ESs of just 0.37, 0.47 and 0.40, respectively. Other teaching practices, such as setting product goals (ES = 0.76), receiving feedback from an adult (ES = 0.80) or providing peer assistance (ES = 0.89), showed to be quite impactful, but findings were based on a reduced sample of studies (between 4 and 7 studies for each instructional approach). Efficacy findings from the SRSD approach were based, in contrast, on 14 studies (Graham et al., 2012, p. 885).

Despite the accumulating evidence in favor of the SRSD approach to improving writing skills and outcomes, we know relatively little about whether such improvement is maintained over time. In this paper, we report findings from a longitudinal follow-up of two large cohorts of children who were administered an SRSD intervention when they were attending 2nd and 4th grade.

What is SRSD and evidence of its effectiveness

Self-Regulated Strategy Development (SRSD) for writing instruction was introduced in the 1980s by Karen Harris (Graham & Harris, 2018) and is firmly grounded in a process-oriented approach to writing (Graham & Harris, 1996; Hayes & Flower, 1980). As such, SRSD interventions typically focus on specific stages of the writing process—such as planning or revising—by offering explicit, easy-to-remember strategies that are straightforward to apply. These process-based strategies are most effective when combined with techniques that enhance students' self-regulation abilities. SRSD instruction is also generally tied to a specific discourse genre or communicative purpose, including opinion essays, narratives, or informational texts (e.g., Graham & Perin, 2007; Graham et al., 2012).

A standard SRSD intervention includes six recursive stages that may be adapted and combined within each session to suit instructional goals and student needs. The first stage, *Develop and activate background knowledge*, encourages students

to reflect on their prior understanding of a given genre, such as the structure and purpose of opinion essays. Teachers often guide discussions about audience, organization, and other key text features, focusing on the development of procedural knowledge. In the *Discuss it* stage, students evaluate their own writing abilities and their use of writing and self-regulation strategies. The *Model it* stage involves the teacher demonstrating the use of target strategies through a "think-aloud" process, openly expressing challenges and showing how the strategies can help resolve them. This modeling also includes the teacher conveying a positive stance toward writing. During the *Memorize it* stage, mnemonic devices are introduced early and consistently reinforced to help students retain the strategies. The *Support it* stage entails active teacher support, which may include guiding questions, progress-monitoring tools, and self-assessment materials. In the final stage, *Independent performance*, support is gradually withdrawn as students become more capable of applying both writing and self-regulation strategies independently.

In summary, SRSD writing instruction involves the explicit teaching of genre-specific knowledge—such as text structures—and the strategic, scaffolded instruction of writing processes, like planning and revising. Instruction is rich in dialogue and integrates direct teaching of self-regulation skills, enabling students to manage the complex tasks involved in writing. The teacher plays a key role in fostering independence, encouraging transfer and maintenance of skills, and promoting positive attitudes toward writing (e.g., Graham & Harris, 1996, 2018; Harris & Graham, 2009).

SRSD has been found to improve children's overall text quality, the number of structurally-relevant discursive elements, as well as other features of text composition. Research shows it performs well across multiple genres (e.g., narrative, argumentative, and informational writing), settings (e.g., individual, small group, or whole class instruction; de la Paz & Graham, 2002; Graham et al., 2012; Rogers & Graham, 2008), and languages (e.g., Portuguese, Limpo & Alves, 2013; Salas, Birello, & Ribas, 2021; Salas, Pascual, Birello, & Cros, 2023; Spanish, Torrance et al., 2007). It has been particularly beneficial for students with learning and language difficulties (Graham & Harris, 2003), as well as for children from disadvantaged backgrounds (e.g., Graham & Harris, 2005; Salas et al., 2021, 2023).

Previous longitudinal studies on SRSD efficacy

SRSD instruction is particularly geared towards maintenance of its effects (Harris & Graham, 1996). Evidence of its immediate effects is quite abundant (Graham & Harris, 2003), while several studies have examined somewhat delayed maintenance. Most of these studies have included relatively small samples of English-speaking participants, often composed of struggling writers. In general, the findings reported point to students successfully maintaining the effects of the SRSD intervention, with similar effect sizes at posttest and maintenance (i.e., the strength of the beneficial effect was maintained for the duration of the study), which generally took place a few weeks after the intervention had concluded (e.g., Graham et al. 2005; Harris et al., 2006; Tracy et al., 2009).

Hacker et al. (2015) assessed posttest and maintenance effects of SRSD for writing persuasive essays on a large sample of 628 7th grade students, quasi-randomly assigned to the SRSD or the control conditions. Although the sample was quite large, it was only drawn from two schools: one where the SRSD intervention took place, and a control school. Students produced three writing samples at pretest, posttest, and maintenance, with the latter taking place about two months after the intervention. Writing samples were scored on six text-quality traits (voice, organization, ideas, conventions, word choice, and sentence fluency), using an automated essay scoring system. Contrary to their initial prediction, both groups showed improvement in writing performance from pre- to posttest, with no significant differences between the two groups. Although the groups did not differ at posttest, the scores of students in the SRSD group were significantly higher than those of students in the control group at maintenance.

A few studies examining maintenance effects of SRSD have been conducted in languages other than English. Festas et al. (2015) examined the effects of SRSD for opinion essay writing on 380 eighth-grade Portuguese students. Pretest, posttest, and maintenance (two months after post-intervention) writing probes were used to assess the effects of SRSD instruction on students' opinion-essay writing. Their results showed that students who received SRSD instruction included more structural elements in their essays at posttest, relative to pretest. Moreover, the improvement was maintained two months after the intervention had concluded. Also, Torrance et al. (2007) investigated the maintenance effects of Cognitive Self-Regulation Instruction (CSRI) to 95 Spanish-speaking 6th graders. This teaching program provides strategies for planning and revising opinion essays, which the authors acknowledged to be similar to SRSD (p. 269). The study assessed maintenance effects of CSRI instruction to write compare-and-contrast texts on both process and product measures. Process measures included computing the total time spent writing, as well as time in specific writing tasks (e.g., planning, translating, revising). Product measures included holistic text quality, coherence, and text structure. In addition, authors assessed text-based measures, including number of words or paragraphs and several linguistic features (e.g., connectivity devices). Results showed that, 12 weeks after the intervention had concluded, students in the CSRI group spent more time composing text than the control group. Most of this extra time was due to more time dedicated, particularly, to planning activities (e.g., thinking about content). In addition, the texts of the CSRI group produced at the delayed posttest were of significantly higher quality, more coherent, and included more discursively-relevant elements (e.g., an introduction or a conclusion).

Also in Spanish, Fidalgo et al. (2008) conducted a follow-up study of Torrance et al. (2007) participants after 28 months had passed. The final sample were 77 students, now attending 8th grade. The authors kept track of the treatment-group students and they recruited controls among their new peers. As a result, there were no baseline measures to compare results to, and the entire study was based on differences between a group of students that received CSRI training two years prior and another group who did not. Students were asked to write a comparative essay on the topic *Spain in the middle ages vs. the Spain of today,* using a 3-page reference text. Results indicated that the texts produced by students in the CSRI condition were of

significantly higher quality, more coherent, and included more relevant structural elements than those produced by students in the control condition. In short, the Fidalgo et al. (2008) study showed that process and product improvements were still detectable 28 months after the intervention.

To conclude, the studies that have been carried out until today examining maintenance of SRSD or SRSD-like writing interventions point clearly to these effects being sustained over time. However, there are some important methodological limitations: several studies included a very limited sample, and all but one covered a relatively short period after the intervention had elapsed. Moreover, there are more studies with participants who attended the last years of primary or early secondary-school years, whereas little data is available on younger participants.

This study

We present follow-up data on a large sample of 2nd and 4th graders, whose schools were randomly allocated to an SRSD opinion-essay planning condition or to a business-as-usual (BAU) control condition. The immediate posttest efficacy of the intervention has already been reported in Salas et al. (2021) and (2023), showing that moderate to large effect sizes were obtained for planning quality, number of structural elements (thesis statement, reasons, explanations, conclusion), and overall text quality. Here, we include three additional time points: a delayed posttest, obtained one month after the intervention; a mid-term follow-up, obtained six months after the intervention; and a long-term follow-up, obtained 18 months after the intervention. That is, we investigated a total of five time points. The intervention was not continued after posttest.

This paper improves and expands on extant literature in at least four ways: (1) we included several writing probes, covering a large time span, from one month to 18 months after the end of the intervention; (2) our sample size is considerably larger than most of those included in previous studies, but we have nevertheless included multiple outcome measures; (3) it involved children at an earlier stage in writing development; specifically, 2nd and 4th grade children, who were moreover compared to each other; and (4) we compared follow-up time points to our baseline measure (i.e., pretest), but also tested differences between adjacent time points (i.e., posttest to delayed posttest; delayed posttest to mid-term follow-up, and so on.).

Our working research questions were as follows,

RQ1: Were the effects of SRSD maintained over time for planning, number of structural elements, and text quality?

RQ2: Was the magnitude of the effects maintained over time?

RQ3: Were any effects affected by participants' level of education (i.e., grade)?

With regards to RQ1, we hypothesized that, in line with previous studies, effects of the SRSD intervention would be retained after posttest for all measures. With regards to RQ2, we expected that the magnitude of effects would decrease over time, with performance scores becoming progressively smaller, particularly at the later follow-ups. Finally, we were ambivalent regarding RQ3, given the absence of previous research on the topic. However, differences as a function of grade in participants'

retention of SRSD-derived strategies seem unlikely, mostly because our previous studies showed virtually no effects of grade in any measure (Salas et al., 2021, 2023). Additionally, Graham and Perin (2007) reported that grade was not a significant moderator of the effect sizes of the writing interventions included in their meta-analysis (p. 464).

Method

Participants

A total of 909 2nd (50.39%) and 4th (49.61%) students participated in this follow-up study. They had been recruited from 12 public schools in [information withheld for the review process]. Schools (and not children or classrooms) had been randomly assigned to one of three conditions: an SRSD opinion-essay planning intervention, an identical SRSD intervention without explicit instruction on the use of connectivity devices (e.g., discourse markers, such as "First", "In conclusion", etc.), or a business-as-usual (BAU) control condition. Salas et al. (2023) found only differences of magnitude, but not in the pattern of results between the two experimental conditions; that is, both conditions showed similar differences with the control group and there were very few significant differences between them. As a result, children in both conditions were collapsed into a single SRSD condition for the present study. Full demographic data is shown on Table 1. Attrition oscillated between 15 and 23% throughout timepoints and groups of children. Note that some classrooms could not implement the delayed posttest because the implementation of the intervention had finished only a few days before the end of the school year. SES levels were estimated using Ganzeboom et al. (1992) equivalence of professions to an SES index, in which higher numbers indicate higher SES levels. For example, a construction worker receives a

Table 1 Full demographic characteristics of the sample

	BAU control		SRSD	
	Grade 2	Grade 4	Grade 2	Grade 4
N at T1	143	118	315	333
(boys, %)	(71, 49.65%)	(53, 44.92%)	(155, 49.21%)	(194, 58.26%)
N at T2, T3, T4 and T5	125, 82, 122, 110	109, 84, 105, 98	278, 185, 268, 265	311, 204, 298, 283
Age at pretest (SD)	7;6 (4.59)	9;5 (4.70)	7;4 (4.41)	9;4 (4.36)
SES	46.85 (15.53)	46.64 (13.43)	48.96 (14.97)	48.79 (15.37)
Min–Max SES	23–86	27–75	23–80	23–79
Exposure to Catalan outside of school (range 0–2.5)	0.54 (0.79)	0.71 (0.96)	0.99 (1.06)	1.00 (1.01)

BAU = Business as usual; SRSD = Self-regulated strategy development; SES = socioeconomic status; T1 = pretest; T2 = posttest; T3 = delayed posttest; T4 = mid-term follow up (6 months); T5 = long-term follow-up (18 months)

score of 32, a real-estate agent, 61, and a physician or dentist, 88. Familiar SES was comparable across conditions, t (770) = − 1.81, p = .072.

There were two main reasons for choosing to randomize at the school level. First, we intended to test the efficacy of SRSD in intact classrooms; therefore, we could not assign individual students to conditions. Second, if we had opted for randomly assigning classrooms to conditions, we would have forced teachers in schools with more than one classroom per grade (the majority of schools) to stop collaborating with each other preparing lesson plans and coordinating other educational activities as they normally do. Also, we ensured that there was no "contamination" across classrooms assigned to different conditions. Another advantage of our procedure was that we were able to offer professional development training not just to the 2nd and 4th grade teachers, but to the rest of the staff, should they be interested. This last aspect of the design was very highly valued by the teachers and, particularly, by the leadership team of the schools, who were often concerned by the suboptimal results of their approach to writing instruction across grades. Consequently, our randomization at the school level facilitated the recruitment of schools to a large degree.

Catalan is the language of instruction in the region where the data were collected. It is a language that coexists with Spanish, with virtually all school-aged children being at least bilingual. Catalan is similar to Spanish both etymologically and typologically. Catalan has, however, a spelling system that is considerably more complex and inconsistent than Spanish (Llauradó & Tolchinsky, 2016; Salas, 2020, 2023). In addition, several children from immigrant origin may not have any exposure to Catalan outside of the school setting. Therefore, we measured the level of exposure to Catalan outside of the classroom, by asking which language(s) children typically used to talk to their mother and father (up to 1 point for each parent, if Catalan was the sole language of communication), to their siblings and other close relatives (up to 0.25 points if Catalan was one of the languages used), and to their friends (up to 0.25 points if Catalan was one of the languages used). The resulting variable ranged, thus, from 0 to 2.5. Children across conditions showed different levels of exposure to Catalan, t(846) = − 4.83, p < .001, indicating that children in the SRSD condition tended to have significantly more exposure. For this reason, we included this variable as a covariate across analyses.

Ethical considerations

The study was approved by the Ethics Committee of the [reference withheld for the review process]. The families or legal tutors of all participants received detailed written information about the purpose, nature, and duration of the study, and signed the corresponding consent forms. Following the Ethics Committee guidelines, parents were knowledgeable that their children could withdraw from the study at any point and with no consequences. We had no withdrawals due to parents' deciding to stop their children participating in the study.

The study involved collecting data at several points during the first year, and again once at the second and third years. The timing of these data collections was negotiated with the schools, to ensure minimal disruptions. The most delicate aspect of the

research was, however, the 11-session intervention on writing, which meant removing something from the regular teaching program for 2nd and 4th grade students at each school. To ensure the least interference with each school's plan and to make sure that children received the full range of subjects expected for their grade, we proceeded as follows: (1) schools were recruited the school year prior to the implementation of the data collection and intervention, so that they could reserve the necessary dates in the 2nd and 3rd trimester for the intervention; (2) we conducted the random allocation of schools to conditions during the first weeks of the new school year, to leave teachers sufficient time to rearrange subjects accordingly; and (3) we made recommendations tailored to each school's specificities as to the best way to schedule the intervention sessions, to make sure that the timing and frequency of the intervention sessions were equivalent across schools.

Writing samples and text-based measures

Children completed a writing task at each time point under the same conditions. The administrator presented the activity, which involved producing a draft and a final text as a response to an opinion-essay prompt. Prompts were counterbalanced at each time point. Table 2 shows the prompts used at each data-collection period. Children were instructed to complete the draft in a maximum of 10 min and to complete the text in up to 20 min. The writing activity was the only task performed in the sessions, which took place in children's regular classrooms. Children were encouraged to write as best as they could, but were offered no help with vocabulary, ideas, or spelling. Several measures were obtained from this task, described below.

Planing quality

Children's drafts were assessed using a modified version of the 1-to-5 scoring system developed by Whitaker et al. (1994). The lowest score, 1, was assigned to plans that were either blank or consisted of text later copied directly into the final version. Extremely brief plans—such as a single sentence, just the prompt, or a single word or phrase—also received a score of 1. Plans demonstrating minimal planning, such as slight differences between the plan and the final text (e.g., small additions, brief elaborations, or minor reordering of ideas), were given a score of 2. A score of 3 was assigned to plans composed primarily of keyword lists. Plans that began to

Table 2 Writing prompts counterbalanced at each time point

Pretest & posttest	Do you think all children your age should go to school? Do you think that having a recess time is necessary?
Delayed posttest	Do you think children should give their toys to children who don't have any? Do you think children your age should have a cellphone?
Mid-term & long-term follow-up	Do you think children should exercise every day? Do you think children your age should go to bed early?

show elaboration of ideas, particularly those indicating structural intent (e.g., labeling where ideas would appear in the text, such as in the introduction, body, or through pros/cons lists), were rated as 4. The highest score, 5, was reserved for plans with clearly elaborated and hierarchically organized ideas. A trained research assistant rated all plans, with a second rater independently scoring 10% of the sample. Interrater reliability was high, with a weighted Kappa of .871.

Number of structural elements

We examined whether children incorporated genre-appropriate structural elements in their essays. Texts were uploaded to a custom-built web application called *TextHandler* (see Salas et al., 2021, for a detailed description of the app). Raters were advanced speech-therapy students or primary school teacher trainees. The app displayed a randomly selected opinion essay, which they assessed for the presence of five key structural elements: introduction, thesis statement, reasons, explanations, and conclusion. Each element present earned one point, and students received additional points for each unique, additional reason and explanation they included in their essay. The total score represented the sum of all points, resulting in a continuous, open-ended measure. Raters had access only to the essay prompt and no other contextual information. They received training via a comprehensive manual detailing the criteria for identifying each structural component. For example, they were instructed to confirm the presence of an explicit thesis statement (excluding vague assertions like "I (don't) think so"), to identify supporting reasons, explanations that developed those reasons (e.g., through examples), and to distinguish between summary or extended conclusions. Raters were told to evaluate the structure rather than the quality or persuasiveness of the arguments, and to award points for any coherent content. Reasons or explanations that were repeated, however, were not to be credited. Each rater initially evaluated 40 texts previously scored by the research team, and proceeded with full scoring only after achieving an intra-class correlation coefficient (ICC) of .80 or above. Two raters assessed all essays, and a third re-evaluated a random 30% of the pre-test and post-test texts. Final interrater reliability was high, with an ICC of .908.

Text quality

TextHandler was also employed to assess the overall quality of the essays using a holistic scoring approach. Raters were trained using a set of 40 benchmark texts from a specific grade level, which had been previously scored by the research team. Similarly to the Number of Structural Elements measure, they proceeded with full scoring only after achieving an intra-class correlation coefficient (ICC) of .80 or above. Upon logging into TextHandler, raters selected the grade level they intended to evaluate (i.e., either 2nd or 4th grade). The app then displayed a randomly-selected essay on the left side of the screen, while a rating scale appeared on the right (see Salas et al., 2021, for a detailed description of the app). Raters assigned a score ranging from 0 (well below grade-level expectations) to 5 (well above grade-level expectations), with a score of 2 representing texts that were considered to be at grade-level expecta-

tions. Each essay was evaluated independently by two raters. If their scores differed by more than one point, a third rater reviewed the essay. Interrater reliability, measured using intra-class correlation coefficients (ICCs) was computed for each grade and time point separately. ICCs for grade 2 ranged between 0.81 and 0.85, while for grade 4 they ranged from 0.80 to 0.85. The final score for each essay was calculated as the average of all ratings.

The interventions

SRSD condition

Children in the SRSD groups participated in 11 one-hour sessions scheduled during regular language instruction time. The intervention was largely based on the intervention developed by Limpo and Alves (2013), which was, in turn, based on the TREE strategy by Harris et al. (2002). Intervention sessions were delivered once or twice per week by the children's regular classroom teachers. Students were informed that they were taking part in a project designed to gather their opinions on meaningful topics, and that they would publish their essays on a blog. Both intervention groups were taught a planning strategy: 2nd grade students learned the "PER" strategy, while 4th grade students learned the "CREC" strategy. Both SRSD groups (i.e., with or without an explicit linguistic instruction) received largely the same intervention; the only difference being that one SRSD group engaged in a few activities that emphasized the use of connectivity devices in opinion-essay writing. Full details are included in Salas et al. (2021, 2023).

Throughout the intervention, teachers were expected to: (1) foster student participation, (2) keep students focused on learning the strategies and ignore concerns like spelling, (3) provide feedback specifically on the use of the taught strategy rather than on mechanical aspects such as punctuation or grammar, and (4) promote students' independent use of the strategies. Unlike many SRSD-based writing programs, this intervention followed a time-based schedule rather than a mastery-based one (Harris & Graham, 2009). Nevertheless, teachers were encouraged to support students in tailoring their goals according to individual ability. For instance, students who struggled could begin by setting simpler goals, such as writing just a thesis statement and one supporting reason, before progressing to more complete plans. Similarly, while students were generally expected to stop using graphic organizers after session eight, continued use was permitted for those who benefited from additional support.

Development of the SRSD intervention Lesson 1 involved introducing students to the project (writing opinions for a blog) and discussing key terminology (e.g., opinion, planning, strategy, goal). Lesson 2 was devoted to presenting the strategy (PER, for 2nd graders, and CREC, for 4th graders). Goal-setting, self-evaluation, and progress monitoring were self-regulation strategies introduced in Lesson 3. Children first evaluated a model text by a "child who had already been taught the PER/CREC strategy for planning opinion essays" and they later self-evaluated their own pretest text on the same topic. Students were led to conclude that the planning strategy was useful for writing good opinion essays. Children in the full SRSD condition

also completed tasks to become aware of the connectors used in the model text and to assess their presence in their own initial texts. In Lesson 4, children in the full SRSD condition carried out a series of activities with connectors, such as pairing a specific connector (e.g., In my opinion…) with a structural element of the text (e.g., Thesis statement). Classrooms in the SRSD condition without connectivity instruction (henceforth, SRSD-), the teacher modeled the use of the strategy to plan an opinion essay (with or without making explicit reference to connectivity devices). Afterwards, the class reflected on the self-instructions that the teacher had used to remind herself of the main steps in producing a plan (i.e., setting a goal, planning with the strategy, and self-evaluating). In Lesson 5, students in the SRSD condition continued carrying out activities on text connectivity, whereas children in the SRSD- condition planned an essay with the supervision of the teacher. In Lesson 6, children in the full SRSD condition saw the teacher's modelling of the strategy and reflected on self-instructions, whereas children in the SRSD- condition planned a text individually. In Lesson 7, children in the SRSD condition planned a text collectively, with the supervision of the teacher, whereas children in the SRSD- condition were modelled the strategy without the use of a graphic organizer. In Lesson 8, both groups planned and wrote a text in pairs without a graphic organizer. In Lesson 9 all students planned and wrote a text individually. Lesson 10 was used, in both groups, to provide feedback to students. Priority was given to problems with the use of the strategy, then to issues using the plans to write the text and, only if children showed mastery of those skills, they were advised on how to improve their texts (e.g., introducing more reasons or developing explanations further). Finally, in Lesson 11 children across groups planned and wrote a text individually.

BAU control condition

Teachers in the control condition were asked to keep diaries documenting any literacy-related activities conducted between the pretest and posttest period. These diaries were reviewed by the first author to confirm that none of the recorded activities overlapped with those used in the experimental condition. Most literacy activities in the control condition focused on foundational skills such as spelling, punctuation, and decoding, along with reading short texts and occasional free writing exercises. To account for practice effects, control condition teachers instructed students to produce the same number of texts using the same prompts as those given to students in the intervention condition. After the study concluded, control-condition teachers received the same training during the following academic year. To the best of our knowledge, no control school implemented the SRSD sessions (either the full version or the one without explicit connectivity instruction) thoroughly or systematically.

Treatment fidelity

Several measures were implemented to ensure treatment fidelity. First, teachers in the experimental condition participated in an 8-h training session at their respective schools, led by either the first author or an experienced continuing-education trainer.

The research team developed detailed teacher manuals outlining all 11 sessions, including step-by-step lesson plans, scripted activities (e.g., for modeling sessions), and answer keys. These manuals were thoroughly reviewed during training. Second, each session included a checklist of essential instructional components, which teachers completed after each lesson to monitor adherence to the planned content. Third, teachers were provided with audio recorders and instructed to record all sessions, with special emphasis on those requiring high teacher involvement (i.e., Sessions 1–7). An external research assistant reviewed a random selection of 30% of these recordings, independently completing the same checklists. This analysis indicated an average lesson coverage rate of 95.14% (SD = 6.66). Finally, the first author and the trainer communicated periodically with the experimental-condition teachers, checking in before and after each session to address any issues from the previous lesson and preview the objectives of the upcoming one.

Analytical strategy

Prior to analyzing the data, we looked at sources of variation at three potential levels: The class level, the individual student level, and the measurement level, by means of intraclass correlation coefficient (ICC); that is, we assessed the proportion of class variation (L3) and individual variation (L2) over the total variation. We learned that both L3 and L2 were un-ignorable. Specifically, our findings revealed that in both upper levels ICCs exceeded the common threshold of 5% (ICC > .05) and even exceeded 20% for some outcome measurements. Thus, a three-level framework was developed to assess intervention and time differences. Specifically, we used the multilevel MIXED model (SPSS V.30) to include the random intercepts for individuals nested within classes. Time (pretest, posttest, delayed posttest, mid-term follow-up, and long-term follow-up), Condition (SRSD vs. BAU control), and Grade (2nd vs. 4th) were entered as fixed effects. Because exposure to Catalan differed between the SRSD and control conditions, we controlled for this variable across analyses. The *F* statistic was used to test fixed effects on each dependent variable: planning quality (PLAN), number of structural elements (STR), and text quality (TQ).

A common practice in regression models, regardless of their type, is to test for main effects first, and to add interaction terms in subsequent steps (Petrocelli, 2003; Schafer, 1991). Therefore, we first ran main effect MIXED models. In a second stage, we included all 2-way interactions. Finally, we included a 3-way interaction. Models were compared using both Akaike's Information Criterion (AIC) and Schwarz's Bayesian Criterion (BIC), with lower values indicating better model fit. Marginal mean (predicted) rankings were based on pairwise multiple comparisons (*p*-values were adjusted by the Bonferroni correction for multiple comparisons).

Table 3 Descriptive statistics of all outcome measures across time points

	Pretest		Posttest		Delayed posttest		Mid-term follow up		Long-term follow-up	
	SRSD *M* (SD)	Control *M* (SD)	SRSD *M* (SD)	Control *M* (SD)	SRSD *M* (SD)	Control *M* (SD)	SRSD *M* (SD)	Control *M* (SD)	SRSD *M* (SD)	Control *M* (SD)
PLAN										
G2	1.34 (0.49)	1.24 (0.45)	4.30 (1.31)	1.19 (0.54)	4.58 (1.05)	1.43 (0.57)	2.23 (1.26)	1.59 (0.92)	1.73 (1.21)	1.20 (0.49)
G4	1.64 (0.84)	1.43 (0.71)	4.06 (1.50)	1.53 (0.92)	3.87 (1.58)	1.62 (0.83)	2.50 (1.31)	1.71 (0.84)	2.22 (1.30)	1.64 (0.92)
STR										
G2	1.69 (1.08)	1.54 (1.10)	3.68 (1.30)	1.66 (0.91)	1.98 (1.93)	0.94 (1.03)	1.97 (1.44)	1.59 (1.31)	1.99 (1.44)	1.42 (1.23)
G4	2.35 (1.38)	2.47 (1.43)	5.64 (2.05)	2.69 (1.19)	3.05 (2.87)	1.51 (1.35)	2.64 (1.37)	2.32 (1.45)	2.77 (1.63)	2.26 (1.46)
TQ										
G2	0.74 (0.60)	0.64 (0.60)	2.52 (1.38)	0.74 (0.58)	2.17 (1.11)	1.00 (0.64)	1.77 (0.93)	1.13 (0.74)	1.14 (0.78)	0.89 (0.59)
G4	0.74 (0.57)	0.82 (0.66)	2.34 (1.14)	0.90 (0.52)	2.16 (1.08)	1.20 (0.69)	1.68 (1.00)	1.03 (0.76)	2.03 (1.02)	1.82 (1.21)

PLAN = Planning quality (range: 1–5); STR = number of structural elements; TQ = text quality (range: 0–5); G2 = Grade 2; G4 = Grade 4

Results

Main effects of condition, grade, and time on performance outcomes

Descriptive statistics are shown in Table 3, while Table 4 provides complete modeling results for main effects, two-, and three-way interactions. A GEE analysis on PLAN scores revealed a significant main effect of Time, $F(4) = 434.81$, $p < .001$, indicating that children improved their planning skills over time. Children across conditions and grades did minimal planning at pretest (around 1.5 out of a maximum of 5 points). In contrast, children in the SRSD groups showed average scores of 4 and above at posttest and the delayed posttest (Table 3). There was also a significant main effect of Condition, $F(1) = 89.32$, $p < .001$, indicating that children in the SRSD condition outscored those in the control condition. However, the effect of Grade was non-significant, $F(1) = 0.57$, $p = .457$, suggesting that 2nd graders and 4th graders showed similar performances. Time differences are shown in Fig. 1. Posttest and delayed posttest scores were the highest and did not differ from one another (d). Pretest scores were the lowest (a), while mid-term follow-up scores (c) were lower than either posttest or delayed posttest scores, but higher than long-term follow-up scores (b).

When we analyzed the number of structural elements in children's opinion essays, GEE analyses showed a significant main effect of Condition, $F(1) = 24.78$, $p < .001$, whereby the SRSD condition outperformed the control condition. Note that 2nd grade children included less that 2 structural elements in their pretest essays, but then included more than 3, on average, at posttest. Fourth graders went from about 2 struc-

Table 4 Main and interaction effects in explaining performance outcomes

Main effects	Planning quality (1–5)	No. of structural elements	Text quality (0–5)
Grade (*F*)	0.57 ^{ns}	36.95***	4.37 ^{ns}
Marginal Means			
Second Graders	2.10^{a}(0.09)	1.88^{a} (0.12)	1.27^{a} (0.06)
Fourth Graders	2.19^{a}(0.09)	2.83^{b} (0.12)	1.45^{b} (0.06)
Condition (*F*)	89.32***	24.78***	48.66***
Marginal Means			
BAU control	1.51^{a}(0.11)	1.79^{a} (0.05)	1.04^{a} (0.08)
SRSD	2.79^{b}(0.07)	2.64^{b} (0.04)	1.69^{b} (0.05)
Time (*F*)	434.81***	227.77***	262.99***
Marginal Means			
Pretest	1.12^{a}(0.08)	1.79^{a} (0.10)	0.58^{a} (0.05)
Posttest	3.10^{b}(0.08)	3.81^{c} (0.10)	1.84^{b} (0.05)
Delayed posttest	3.14^{c}(0.08)	2.03^{ab} (0.10)	1.67^{c} (0.06)
Mid-term follow up	1.85^{c}(0.08)	2.05^{b} (0.10)	1.38^{d} (0.05)
Long-term follow up	1.52^{c}(0.08)	2.08^{b} (0.10)	1.37^{d} (0.05)
Exposure to Catalan	804.19*	15.67***	21.26***
Effect (regression coefficient b)	2.14*** (0.11)	2.02*** (0.14)	1.64*** (0.76)
AIC	11,431.15	15,638.34	9,515.00
BIC	11,449.67	15,657.30	9,533.55
Two-way interactions	*F*	*F*	*F*
Grade × Condition	0.81 ^{ns}	0.22 ^{ns}	0.945 ^{ns}
Grade × Time	17.09***	16.86***	58.99***
Condition × Time	195.51***	74.42***	119.58***
AIC	10,712.95	15,300.62	8,911.95
BIC	10,731.46	15,319.57	8,930.49
Three-way interaction			
Grade × Condition × Time	4.40**	4.41***	1.18 ^{ns}
AIC	10,700.22	15,285.87	8,914.92
BIC	10,718.73	15,304.82	8,933.45

***$p < .001$, **$p < .01$, *$p < .05$; standard errors of the estimates in parentheses

tural elements at pretest to more than 5 at posttest (Table 3). The difference between 4th graders and 2nd graders was statistically significant, $F(1) = 36.95$, $p < .001$. The mean number of structural elements varied significantly over time, $F(4) = 227.77$, $p < .001$, with posttest scores (c) being the highest (Fig. 2). Later STR scores were significantly lower and did not differ from each other. Pretest levels (a) were lowest and differed significantly from all later scores.

Across the board, Text quality (TQ) scores were below grade-level expectations at pretest (i.e., under 2; Table 3). At posttest and at some points after, children in the SRSD condition score at or above this level. Accordingly, TQ was found to be significantly higher among the SRSD, rather than control-condition children, $F(1) = 48.66$, $p < .001$, and among 4th graders, rather than 2nd graders, $F(1) = 4.37$, $p = .044$.

Time differences were also statistically significant, $F(4) = 262.99$, $p < .001$, shown in Fig. 3. Posttest scores were highest (d) and pretest scores were lowest (a).

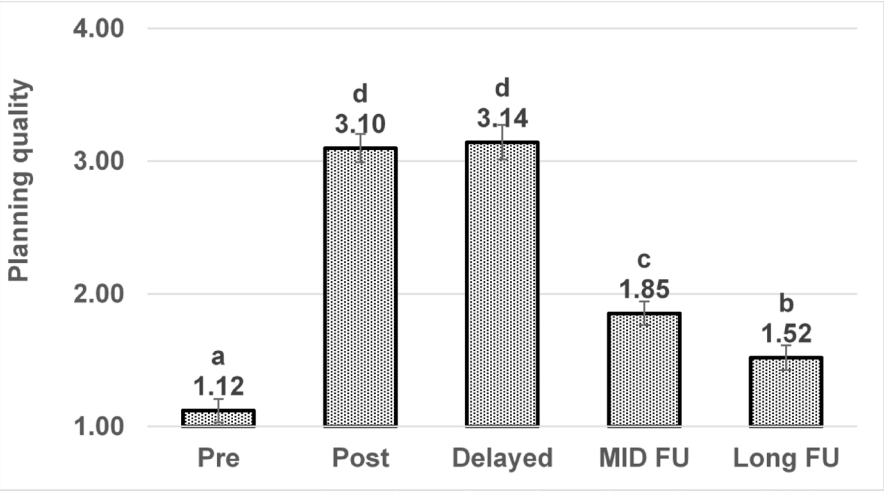

Fig. 1 Time effect for planning quality (PLAN). Latin letters for predicted marginal mean rankings based on multiple pairwise comparisons adjusted by the Bonferroni correction for multiple comparisons

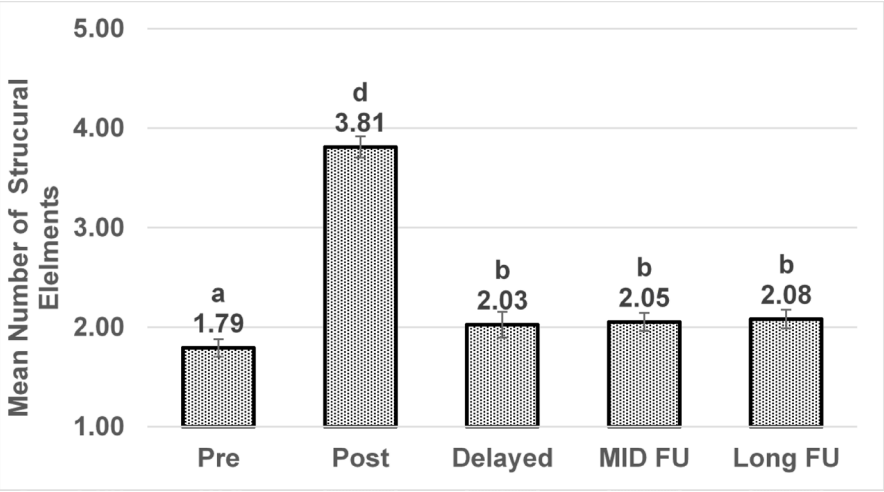

Fig. 2 Time effect for mean number of structural elements (STR). Latin letters for predicted marginal mean ranking based on multiple pairwise comparisons adjusted by the Bonferroni correction for multiple comparisons

The delayed posttest scores were significantly lower than posttest scores but higher than later timepoints. The two latest time points (b) were significantly lower all previous scores, but they did not differ from each other.

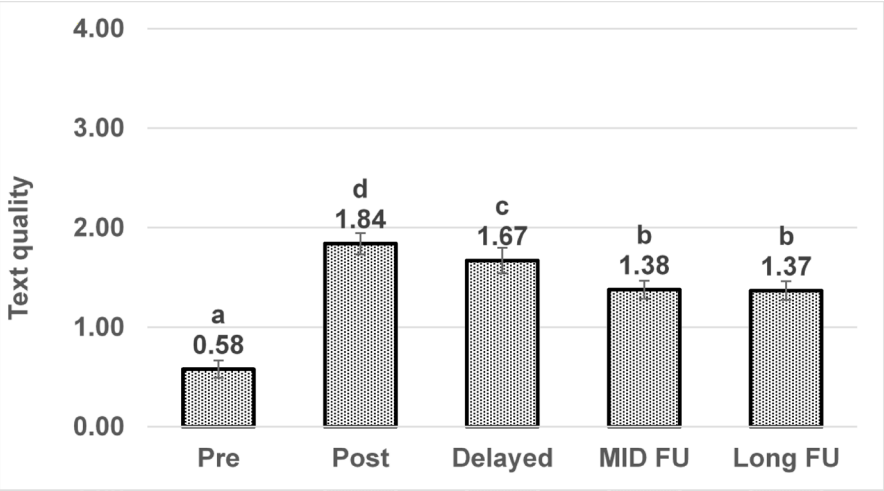

Fig. 3 Time effect for text quality (TQ). Latin letters for predicted marginal mean ranking based on multiple pairwise comparisons adjusted by the Bonferroni correction for multiple comparisons

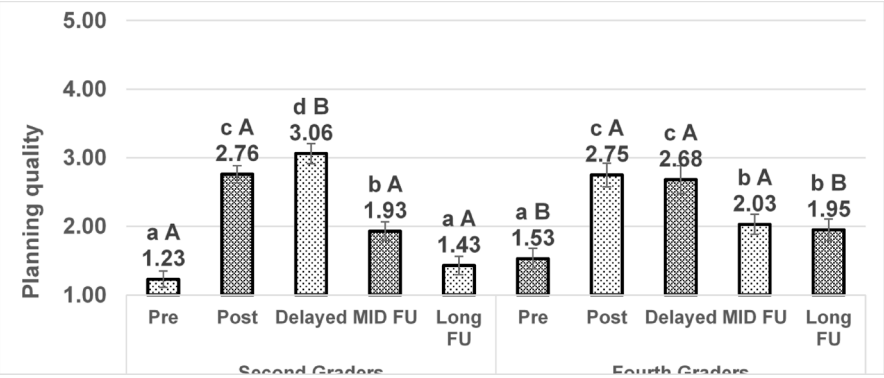

Fig. 4 Two-way interaction analysis between grade and time for PLAN. Small Latin letters for within grade by time ranking; capital Latin letters for within time by grade ranking

Exposure to catalan

Exposure to Catalan outside school was found to be significant in explaining all three performance outcomes, PLAN: $F(1) = 4.22$, $p = .040$; STR: $F(1) = 15.67$, $p < .001$; and TQ: $F(1) = 21.26$, $p < .001$. Higher exposure was associated with higher outcome levels, PLAN: $b = 0.07$ ($p = .040$); STR: $b = 0.18$ ($p < .001$); TQ: $b = 0.15$ ($p < .001$).

Interaction effects of time, grade, and condition on performance outcomes

Planning quality

In our assessment of PLAN, we found grade and condition interactions with time to be significant, $F(4) = 17.09$, $p < .001$; and $F(4) = 195.51$, $p < .001$, respectively. However, the interaction between grade and condition was not significant, $F(1) = 0.81$, $p = .374$.

The Grade by Time interaction effect on PLAN (Fig. 4) showed that, across grades, pretest scores were lowest, and posttest and delayed-posttest scores were highest. Second graders delayed posttest scores were significantly higher than at posttest, and then scores declined first in the mid-term follow-up and worsen even more by the long-term follow-up, the scores of which were not significantly superior to pretest scores.

In 4th grade, children retained the improvements at posttest up to the delayed posttest, but the decline was significant at both follow-up timepoints. Here, however, the long-term scores were superior to the pretest scores. The simple effect of grade at each time point showed that there were no significant differences at posttest and the mid-term follow-up; 4th graders outscored 2nd graders at pretest, whereas 2nd graders significantly outscored 4th graders at the delayed posttest and the long-term follow-up.

The significant interaction between Condition and Time is shown in Fig. 5. Within the BAU control condition, there was little variation in PLAN over time. Within the SRSD condition, PLAN scores peaked at posttest and were maintained at the delayed posttest. There was a significant decline from delayed posttest to the mid-term follow-up, and again another decline at the long-term follow-up. However, even 18 months after the intervention had ceased, children's scores were superior to pretest levels. As for the simple effect of Condition, there were significant differences favoring the SRSD children from posttest onwards.

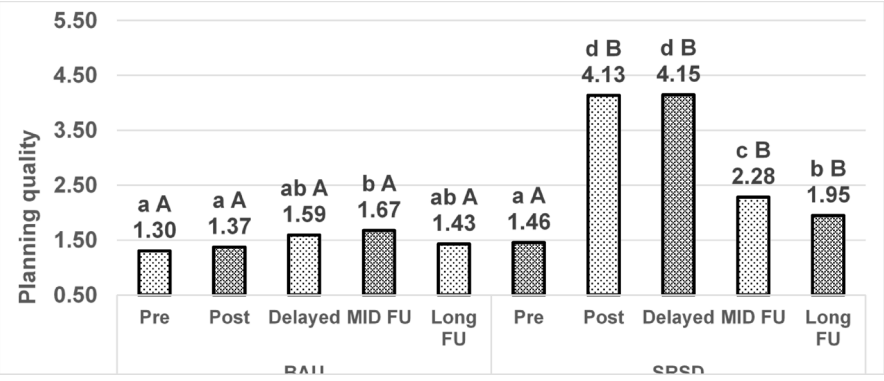

Fig. 5 Two-way interaction analysis between condition and time for PLAN. Small Latin letters for within grade by time ranking; capital Latin letters for within time by grade ranking

Table 5 Predicted marginal means for the three-way interaction of Time × Grade × Condition for PLAN

Time	Condition	Grade	EMM* (SE)
Pretest	BAU control	2	1.22 (0.18)
		4	1.39 (0.19)
	SRSD	2	1.32 (0.12)
		4	1.59 (0.11)
Posttest	BAU control	2	1.22 (0.17)
		4	1.54 (0.19)
	SRSD	2	4.24 (0.12)
		4	4.03 (0.11)
Delayed posttest	BAU control	2	1.49 (0.19)
		4	1.65 (0.19)
	SRSD	2	4.48 (0.24)
		4	3.84 (0.11)
Mid-term follow-up	BAU control	2	1.62 (0.18)
		4	1.74 (0.19)
	SRSD	2	2.26 (0.12)
		4	2.30 (0.12)
Long-term follow-up	BAU control	2	1.20 (0.18)
		4	1.67 (0.19)
	SRSD	2	1.71 (0.12)
		4	2.18 (0.11)

* Range: 1–5

PLAN = Planning quality; EMM = Estimated Marginal Means

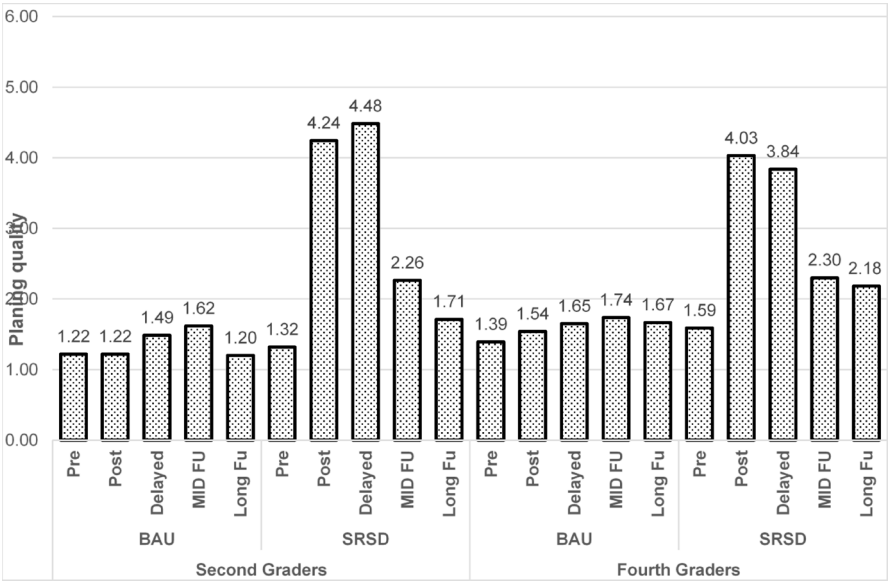

Fig. 6 Three-way interaction Analysis, with predicted marginal means of PLAN

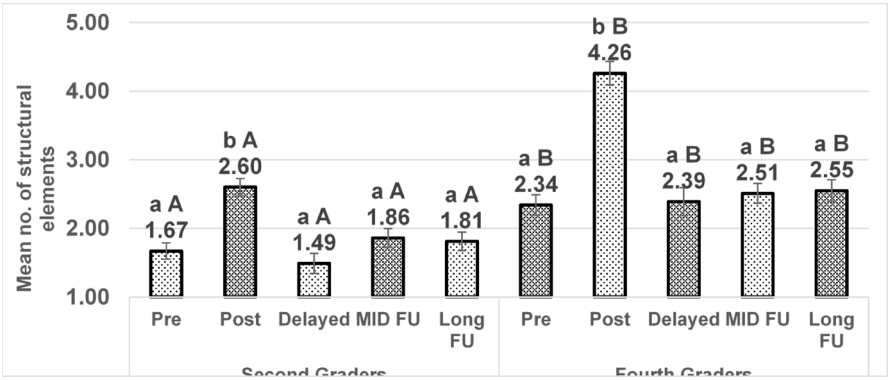

Fig. 7 Two-way interaction analysis between grade and time on STR. Small Latin letters for within grade by time ranking; capital Latin letters for within time by grade ranking

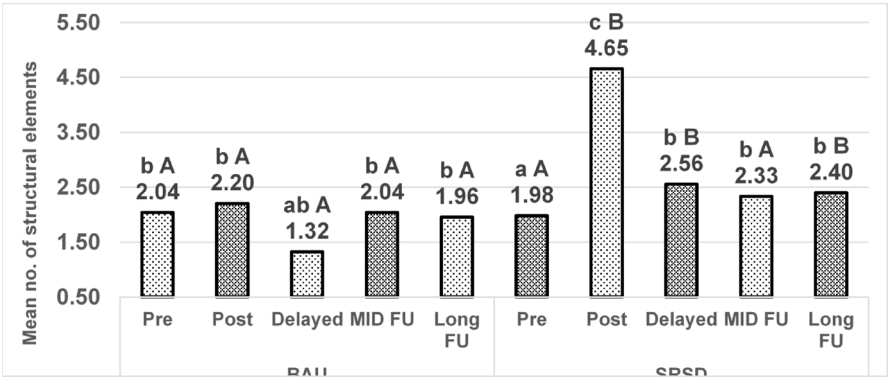

Fig. 8 Two-way interaction analysis between condition and time on STR. Small Latin letters for within condition by time ranking; capital Latin letters for within time by condition ranking

Finally, the Time * Grade * Condition interaction was found to be significant, $F(4) = 4.40$, $p = .002$. Table 5 shows the predicted marginal means of the three-way interaction; that is, subgroup means of time in each condition by grade; see also Fig. 6.

Number of structural elements

The Grade * Time interaction was significant, $F(4) = 16.86$, $p < .002$ (Fig. 7). A breakdown of such interaction revealed that 2nd graders were outscored by 4th graders across time points. Also across grades, only the posttest scores were significantly higher than the rest. An examination of the Time * Condition interaction, $F(4) = 74.42$, $p < .001$ (Fig. 8), indicated that the simple effect of time in the BAU control condition indicated that there was very little variation in STR over time. In contrast, within the SRSD condition, the lowest scores were observed at pretest, and the highest scores were observed at posttest; scores at the delayed posttest and until

Table 6 Predicted Marginal Means for the Three-way Interaction of Time × Grade × Condition for STR

Time	Condition	Grade	EMM (SE)
Pretest	BAU control	2	1.56 (0.24)
		4	2.52 (0.25)
	SRSD	2	1.67 (0.16)
		4	2.29 (0.15)
Posttest	BAU control	2	1.66 (0.24)
		4	2.74 (0.25)
	SRSD	2	3.68 (0.16)
		4	5.62 (0.15)
Delayed posttest	BAU control	2	1.03 (0.24)
		4	1.61 (0.25)
	SRSD	2	2.03 (0.16)
		4	3.08 (0.15)
Mid-term follow-up	BAU control	2	1.65 (0.24)
		4	2.43 (0.25)
	SRSD	2	2.01 (0.16)
		4	2.66 (0.15)
Long-term follow-up	BAU control	2	1.56 (0.24)
		4	2.36 (0.25)
	SRSD	2	2.02 (0.16)
		4	2.79 (0.15)

STR = Mean number of structural elements

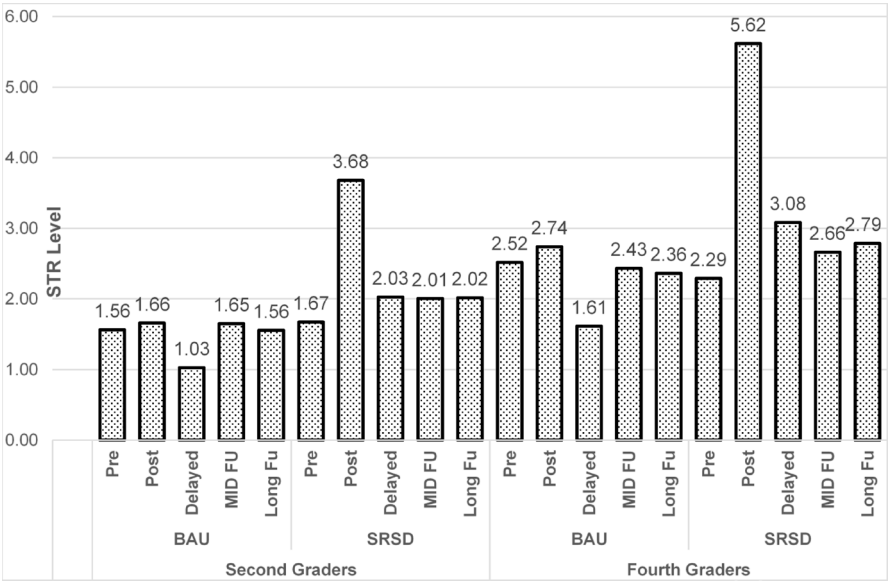

Fig. 9 Three-way interaction analysis, with predicted marginal means of STR

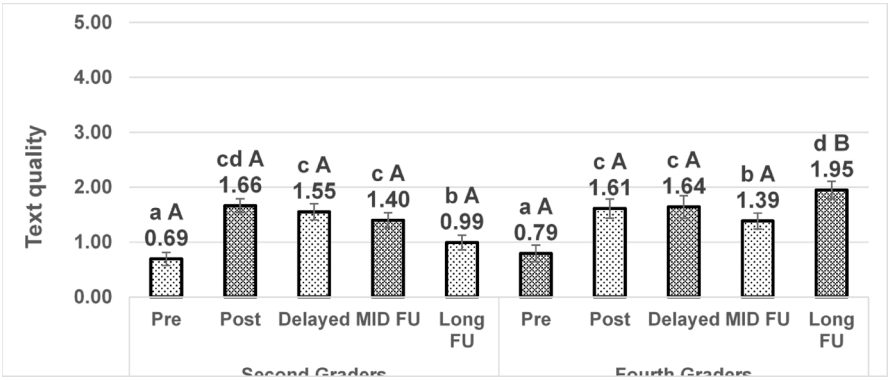

Fig. 10 Two-way interaction analysis between grade and time on TQ**.** Small Latin letters for within grade by time ranking; capital Latin letters for within time by grade ranking

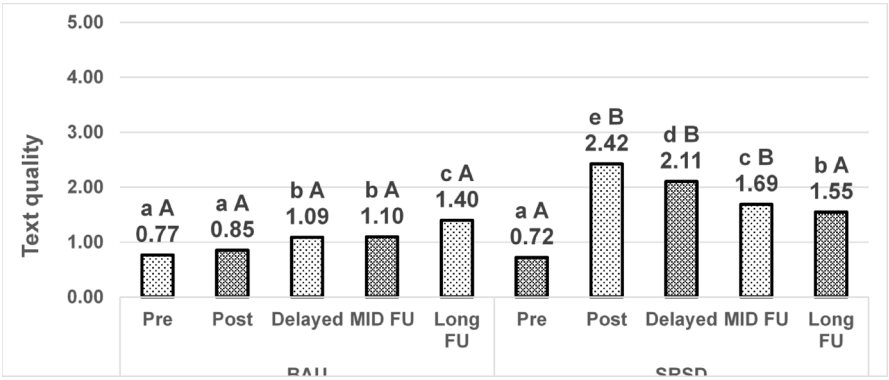

Fig. 11 Two-way interaction analysis between condition and time on TQ. Small Latin letters for within condition by time ranking; capital Latin letters for within time by condition ranking

the end of the study period did not differ from each other and were all significantly lower than posttest scores but higher than pretest scores. Although both conditions included a similar average number of structural elements at posttest, the SRSD significantly outscored the BAU control group at all remaining timepoints, except in the mid-term follow-up. In contrast, the Grade × Condition interaction was not significant, $F(1) = 0.220$, $p = .642$. Finally, the Time * Grade * Condition interaction was found to be significant, $F(4) = 4.41$, $p = .001$. Table 6 shows the predicted marginal means of the three-way interaction; that is, subgroup means of time in each condition by grade; see also Fig. 9.

Text quality

The Grade * Time interaction was significant, $F(4) = 58.99$, $p < .001$ (Fig. 10), and showed that 4th graders only outscored 2nd graders at the long-term follow-up. With regards to the simple effect of Time, the 2nd grade cohort had its lowest TQ

scores at pretest, followed by the long-term follow-up, in turn followed by both the delayed posttest and the mid-term follow-up, which did not differ from each other. Posttest scores were the highest but did not differ significantly from the delayed posttest scores. The 4th grade cohort also obtained its lowest TQ scores at pretest but, in contrast, the highest scores were observed at the long-term follow-up. The posttest and delayed posttest scores came next and did not differ from each other, while the mid-term follow-up showed the second lowest scores.

The Condition * Time interaction, $F(4) = 58.99$, $p < .001$ (Fig. 11) revealed that, within the BAU condition, there was a modest increase in TQ over time: the lowest scores were obtained at pretest and posttest, which did not differ from each other; then a small, significant increase was observed in the delayed posttest and the mid-term follow-up scores, which also did not differ from each other. Last, the highest score was registered at the long-term follow-up. Within the SRSD condition, the lowest scores were observed at pretest and the highest, at posttest. After that, there was a significant decline from one time point to the next. Of note, however, is the fact that the long-term follow-up TQ scores were significantly higher than pretest scores. The simple effect of Condition indicated that higher scores were been obtained by children in the SRSD condition in comparison with children in the BAU control condition at all timepoints, except at pretest and at the long-term follow-up. Finally, the Grade * Condition interaction was not significant, $F(1) = 0.945$, $p = .338$, nor was the triple Grade × Condition × Time interaction, $F(4) = 1.18$, $p = .316$.

Discussion

The effectiveness of SRSD for improving writing skills in the short term is a well-established finding of the science of writing (e.g., Graham et al., 2012, 2025). In this study, we sought to investigate the much less-examined after-posttest effects of SRSD. For this purpose, we followed a relatively large sample of children at the initial (2nd grade) and intermediate (4th grade) stages of learning to write, whose schools had been randomly allocated to an SRSD or to a control condition for learning to plan opinion essays. While we were certain of the immediate, posttest efficacy of the intervention on several writing performance measures (Salas et al., 2021, 2023), we examined whether these effects were retained over a follow-up period of up to 18 months after the end of the intervention.

Our first research question addressed whether the effects were maintained at different points: a delayed posttest, obtained one month after the end of the intervention; a mid-term follow-up, obtained 6 months after the end of the intervention; and a long-term follow-up, obtained 18 months after the end of the intervention. A major first finding of the current study was that the effects of the SRSD opinion-essay planning intervention were, for the most part, long-lasting. Even after 18 months had passed from the end of the intervention, children in the SRSD condition composed better plans and included more structurally-relevant elements in their opinion essays than those of their control-condition peers. Text-quality ratings at the furthest timepoint were superior than at pretest, but the advantage of the SRSD groups over the BAU control students only maintained up to the mid-term follow-up (i.e., 6 months after the

intervention had finished). This finding is in line with previous studies on SRSD writing instruction that also found evidence of maintenance effects in children's writing skills (e.g., Festas et al., 2015; Fidalgo et al., 2008; Graham & Harris, 2003; Hacker et al., 2015). The current study has, thus, contributed to consolidating our understanding of the strong impact of SRSD on writing performance and it has extended it in at least three ways: (1) it has shown that SRSD has long-term effects for children in the early and middle elementary school years; (2) it has provided evidence of the long-term effectiveness of SRSD in Catalan, a language that had not been examined in this way before; and (3) it has overcome several methodological shortcomings of prior studies, such as a limited sample, the absence of true randomization, or a lack of pretest or baseline comparison measures (e.g., Fidalgo et al., 2008; Fitzpatrick & McKeown, 2020; Hacker et al., 2015; Zumbrunn & Bruning, 2023).

Once established that our SRSD writing intervention had long-term effects on children's writing skills, our second research question inquired whether the strength of such effects was sustained or whether it was reduced over time. Our findings indicated that, with time, there was a substantial reduction in the magnitude of the effect of the intervention across writing performance measures. Each outcome measure showed a slightly different pattern: children in the SRSD condition maintained the same levels of planning from posttest to the delayed posttest and progressively declined thereafter. The number of structural elements they included at posttest showed a sharp decline already at the delayed posttest and then plateaued. The quality of their texts experienced a more progressive decline that started at the delayed posttest, with quality significantly dimishing slightly at every successive timepoint. Nevertheless, the effects of SRSD did not completely wash out: even the lowest scores registered after posttest were significantly superior to pretest levels across measures.

Similarly to our own findings, a number of studies have also found evidence of SRSD intervention effects being retained at maintenance. Most often, such probes were administered two to 10 weeks after the end of instruction (e.g., Graham et al., 2005; Harris et al., 2006, 2015; Tracy et al., 2009). However, similar studies have reported intervention effects that were reduced or even cancelled out at maintenance (e.g., Fitzpatrick & McKeown, 2020; Harris et al., 2018; Sawyer et al., 1992). It is somewhat difficult to make a fair comparison between our findings and previous reports of sustained effects of SRSD writing instruction. Some of these studies involved special-needs populations, such as struggling writers (e.g., Graham et al., 2005; Harris et al., 2006; Sawyer et al., 1992). Other studies applied very different methodological designs. In the study by Fidalgo et al. (2008), for instance, there was no baseline measure to compare with outcomes 28 months after the end of the intervention. Yet other studies only compared maintenance scores to pretest levels but not to posttest ones (Festas et al., 2015; Torrance et al., 2007). In the study by Festas et al. (2015) it is possible to see, nevertheless, that there is indeed a decline with time in the average number of structural elements included in children's persuasive texts: from 3.56 at pretest, to 6.07 at posttest, to 4.86 at the delayed posttest (p. 22).

It is interesting to contrast our findings regarding the reduction in intervention effects over time to studies of long-term effects in other (non-SRSD) writing and reading interventions. Hier and Ecker (2016), for example, found that including a

"multiple exemplar training"[1] component to a Performance Feedback (PF) writing intervention used in a previous study –which had shown no maintenance effects (Hier & Eckert, 2014)—increased chances of maintenance at a 4-months follow up. This means that small variations in the delivery of instruction may alter the retention potential and pattern of effects. Interestingly, in a meta-analysis of reading interventions, Suggate (2016) noted that several other factors have an impact on retention of instruction, such as the type of skills or outcomes treated and/or investigated, the target population (i.e., typically-developing vs. learning-disabled), as well as methodological issues (e.g., treatment fidelity or sample size).

Our third research question involved investigating whether the effects of the SRSD intervention would be retained to a different extent by our 2nd and 4th graders. Based mostly on the absence of a moderating effect of Grade in previous studies (Graham & Perin, 2007; Salas et al., 2021, 2023) we hypothesized that both grades would show similar maintenance patterns. Our findings, however, point 2nd graders often retaining effects for less time than 4th graders. This was found for both planning quality and the number of structural elements, but not for text quality, where across grades pretest levels were significantly lower than the latest timepoint. These results do not align well with moderator analyses of Grade in writing instruction (Graham & Perin, 2007), but they do resemble trends detected in a meta-analysis of reading intervention, which suggested that younger children tended to show worse retention rates than older students (Suggate, 2016). Perhaps future research, including metanalyses, should explore the moderating role of educational level on writing intervention effectiveness, differentiating between short-term and long-term outcomes.

Limitations

Longitudinal follow-ups of experimental education studies are rare and for good reasons: they are extremely complex to carry out and it is very difficult to control for the numerous factors that are known to impact outcomes. In the present study, it would have been ideal to systematically obtain detailed knowledge of teachers' practices across conditions during the follow-up period. We did not expect children in the SRSD condition to continue receiving support in the use of the strategies, among other reasons, because their subsequent teachers may not be the ones who received training, since children moved up one grade from the delayed posttest to the mid-term follow-up, and again from the mid- to the long-term follow-up. Anecdotal evidence gathered during the collection of writing samples corroborated this assumption. Conversely, although teachers in the control condition did receive the training after the posttest measures had been collected, we saw no evidence that they were using them systematically. Future attempts at conducting this type of studies should plan for a fine-grained description of literacy practices throughout the study period.

Another limitation of the current study is the assessment of writing skills consisting of a single written production in the target genre. As this study was part of a larger

[1] Multiple exemplar training is defined by the authors as a technique consisting of "varying the training stimuli and/or responses […] providing students with diverse practice" (Hier & Eckert, 2016, p. 113).

project that collected other literacy measures, it was not feasible to ask schools to hold additional testing sessions to obtain multiple samples. We believe that this limitation is partially compensated by the strong within-subject design, which allowed us to obtain several samples of each student throughout the duration of the study. However, the reader should bear in mind that a single writing sample per child provides only a snapshot of their ability at any point in time. Finally, we did not carry out a detailed assessment of which specific structural elements were affected by the intervention, and which ones were lost as a function of time. This crucial aspect of SRSD instruction should be explored further in future studies.

Educational implications and concluding remarks

It is nothing short of outstanding that an 11-session writing program on planning yielded detectable effects across several writing performance measures a year and a half after instruction had concluded. Although teachers did not continue providing support for the strategies, there were observable differences between the control and the SRSD conditions, and in the comparison between baseline and long-term levels. Despite its remarkable efficacy, the benefits obtained from SRSD, as well as from other writing and reading interventions, experiment a sharp decline after only a few months of interrupting instruction. Moreover, our findings point to younger children being more affected by the passing of time than slightly older children. A logical educational implication ensues, such that taught (and learned!) strategies need to be supported in the mid- and the long-term so that students can continue producing high-quality texts. This means that writing instruction needs to carefully consider not just introducing students to effective writing strategies for different communicational settings but also providing them with ample opportunities and support to maintain them over time. Moreover, the finding that 2nd grade children have not just acquired a complex, planning strategy for a discourse genre that is not included in the Catalan curriculum until much later is remarkable. The fact that they were able to maintain their performance level for as long as they did should suffice to warn educators and policymakers of the need to revise the teaching of high-level writing skills at much earlier stages.

To conclude, the present study has presented compelling evidence of the long-lasting impact of SRSD on writing skills in relatively young children; at the same time, however, it has underscored the vital importance of continually and explicitly promoting the use of the strategies, if one wants to avoid the progressive but certain loss of skill.

Acknowledgements We are thankful to the schools and children who participated in the study. We are also indebted to Dr. Gabriel Liberman and Dr. Amir Hefetz, who assisted us with the analytical strategy. This study was conducted thanks to grants PID2019-108791GA-I00 and PID2022- 137868NB-I00, funded by the Ministerio de Ciencia, Innovación y Universidades of Spain, awarded to the first author.

Author contributions N. S. was responsible for the conceptualization, funding acquisition, methodology and data analysis, and co-wrote the original draft, as well as the subsequent revised versions. J. V. and A.

L. conducted part of the literature search and preliminary analyses, co-wrote the original draft, as well as subsequent revised versions. All authors reviewed the manuscript.

Funding Open Access Funding provided by Universitat Autonoma de Barcelona. Ministerio de Ciencia e Innovación, PID2019-108791GA-I00, PID2019-108791GA-I00.

Data availability The anonymized dataset, as well as the scripts for the statistical analyses, are available from the first author upon request.

Declarations

Conflict of interest The authors declare no competing interests.

References

Collins, A., Ciullo, S., Graham, S., & won Lee, J. (2024). Evaluating the effectiveness of writing interventions for students in grades K-5: a comprehensive meta-analysis. *Society for Research on Educational Effectiveness*.

De La Paz, S., & Graham, S. (2002). Explicitly teaching strategies, skills, and knowledge: Writing instruction in middle school classrooms. *Journal of Educational Psychology, 94*(4), 687. https://doi.org/10.1037/0022-0663.94.4.687

Festas, I., Oliveira, A. L., Rebelo, J. A., Damião, M. H., Harris, K., & Graham, S. (2015). Professional development in self-regulated strategy development: effects on the writing performance of eighth grade Portuguese students. *Contemporary Educational Psychology, 40*, 17–27. https://doi.org/10.1016/j.cedpsych.2014.05.004

Fidalgo, R., Torrance, M., & García, J. N. (2008). The long-term effects of strategy- focussed writing instruction for grade six students. *Contemporary Educational Psychology, 33*(4), 672–693. https://doi.org/10.1016/j.cedpsych.2007.09.001

FitzPatrick, E., & McKeown, D. (2020). Writing is revising: Improving student persuasive writing through individualized asynchronous audio feedback. *Education and Treatment of Children, 43*(1), 35–48. https://doi.org/10.1007/s43494-020-00004-4

Ganzeboom, H. B., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research, 21*(1), 1–56. https://doi.org/10.1016/0049-089X(92)90017-B

Graham, S., Cao, Y., Kim, Y. S. G., Lee, J., Tate, T., Collins, P., Cho, M., Moon, Y., Chung, H. Q., & Olson, C. B. (2025). Effective writing instruction for students in grades 6 to 12: A best evidence meta-analysis. *Reading and Writing*, *38*(4), 1–46.

Graham, S., & Harris, K. R. (1996). Self-regulation and strategy instruction for students who find writing and learning challenging. In M. Levy & S. Ransdell (Eds.), *The Science Of Writing: Theories, Methods, Individual Differences, And Applications* (pp. 347–360). Lawrence Erlbaum Associates.

Graham, S., & Harris, K. R. (2003). Students with learning disabilities and the process of writing: A meta-analysis of SRSD studies. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of Learning Disabilities* (pp. 323–344). Guilford Press.

Graham, S., & Harris, K. R. (2005). Improving the writing performance of young struggling writers: Theoretical and programmatic research from the center on accelerating student learning. *The Journal of Special Education, 39*(1), 19–33. https://doi.org/10.1177/00224669050390010301

Graham, S., & Harris, K. R. (2006). Strategy instruction and the teaching of writing. In I. C. A. MacArthur, S. Graham, & J. Fitzpatrick (Eds.), *Handbook of Writing Research.* Guilford Press.

Graham, S., & Harris, K. R. (2018). An examination of the design principles underlying a self-regulated strategy development study. *Journal of Writing Research, 10*(2), 139–171. https://doi.org/10.17239/jowr-2018.10.02.01

Graham, S., Harris, K. R., & Mason, L. H. (2005). Improving the writing performance, knowledge, and self-efficacy of struggling young writers: The effects of self-regulated strategy development. *Contemporary Educational Psychology, 30*(2), 207–241. https://doi.org/10.1016/j.cedpsych.2004.08.001

Graham, S., Kiuhara, S. A., & MacKay, M. (2020). The effects of writing on learning in science, social studies, and mathematics: A meta-analysis. *Review of Educational Research, 90*(2), 179–226. https://doi.org/10.3102/0034654320914744

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

Graham, S., & Perin, D. (2007). Writing next: Effective strategies to improve writing of adolescents in middle and high schools–A report to Carnegie Corporation of New York. Alliance for Excellent Education. https://www.carnegie.org/media/filer_public/3c/f5/3cf58727-34f4-4140-a014-723a00ac56f7/ccny_report_2007_writing.pdf.

Hacker, D. J., Dole, J. A., Ferguson, M., Adamson, S., Roundy, L., & Scarpulla, L. (2015). The short-term and maintenance effects of self-regulated strategy development in writing for middle school students. *Reading & Writing Quarterly, 31*(4), 351–372. https://doi.org/10.1080/10573569.2013.869775

Harris, K. R., & Graham, S. (2009). Self-regulated strategy development in writing: Premises, evolution, and the future. *British Journal of Educational Psychology, 2*(6), 113–135. https://doi.org/10.1348/978185409X422542

Harris, K. R., Graham, S., & Mason, L. (2002). POW plus TREE equals powerful opinion essays. *Teaching Exceptional Children, 34*(5), 74–77. https://doi.org/10.1177/004005990203400513

Harris, K. R., Graham, S., & Mason, L. H. (2006). Improving the writing, knowledge, and motivation of struggling young writers: Effects of self-regulated strategy development with and without peer support. *American Educational Research Journal, 43*(2), 295–340. https://doi.org/10.3102/00028312043002295

Harris, K. R., Graham, S., & Adkins, M. (2015). Practice-based professional development and self-regulated strategy development for Tier 2, at-risk writers in grade 2. *Contemporary Educational Psychology, 40*, 5–16. https://doi.org/10.1016/j.cedpsych.2014.02.003

Harris, K. R., & Graham, S. (1996). *Making the Writing Process Work: Strategies for Composition and Self-Regulation* (2nd ed.). Brookline Books.

Harris, K. R., Graham, S., Mason, L. H., McKeown, D., & Olinghouse, N. (2018). Self-regulated strategy development in writing. In L. Meltzer (Ed.), *Executive Function in Education: From Theory to Practice.* NY: Guilford Press.

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive Processes in Writing* (pp. 3–30). Lawrence Erlbaum Associates.

Hier, B. O., & Eckert, T. L. (2014). Evaluating elementary-aged students' abilities to generalize and maintain fluency gains of a performance feedback writing intervention. *School Psychology Quarterly, 29*(4), 488. https://doi.org/10.1037/spq0000040

Hier, B. O., & Eckert, T. L. (2016). Programming generality into a performance feedback writing intervention: A randomized controlled trial. *Journal of School Psychology, 56*, 111–131. https://doi.org/10.1016/j.jsp.2016.03.003

Limpo, T., & Alves, R. A. (2013). Modeling writing development: Contribution of transcription and self-regulation to Portuguese students' text generation quality. *Journal of Educational Psychology, 105*(2), 401. https://doi.org/10.1037/a0031391

Llauradó, A., & Tolchinsky, L. (2016). The developmental pattern of spelling in Catalan from first to fifth school grade. *Writing Systems Research, 8*(1), 64–83. https://doi.org/10.1080/17586801.2014.1000812

Petrocelli, J. V. (2003). Hierarchical multiple regression in counseling research: Common problems and possible remedies. *Measurement and Evaluation in Counseling and Development, 36*(1), 9–22. https://doi.org/10.1080/07481756.2003.12069076

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

Salas, N. (2020). Non-phonological strategies in spelling development. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2020.01071

Salas, N., Birello, M. & Ribas, T. (2021). Effectiveness of an SRSD writing intervention for low- and high-SES children. *Reading and Writing*, *34*, 1653–16 https://doi.org/10.1007/s11145-020-10103-8

Salas, N., Pascual, M., Birello, M., & Cross, A. (2023). Embedding explicit linguistic instruction in an SRSD writing intervention. *Written Communication*, *40*(3), 857–891. https://doi.org/10.1177/07410883231169516

Sawyer, R. J., Graham, S., & Harris, K. R. (1992). Direct teaching, strategy instruction, and strategy instruction with explicit self-regulation: Effects on the composition skills and self-efficacy of students with learning disabilities. *Journal of Educational Psychology, 84*(3), 340. https://doi.org/10.1037/0022-0663.84.3.340

Schafer, W. D. (1991). Reporting hierarchical regression results. *Measurement and Evaluation in Counseling and Development, 24*(3), 98–100.

Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities, 49*(1), 77–96. https://doi.org/10.1177/0022219414528540

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

Tracy, B., Reid, R., & Graham, S. (2009). Teaching young students strategies for planning and drafting stories: The impact of self-regulated strategy development. *The Journal of Educational Research, 102*(5), 323–332. https://doi.org/10.3200/JOER.102.5.323-332

Whitaker, D., Berninger, V., Johnston, J., & Swanson, H. L. (1994). Intraindividual differences in levels of language in intermediate grade writers: Implications for the translating process. *Learning and Individual Differences, 6*(1), 107–130. https://doi.org/10.1016/1041-6080(94)90016-7

Zumbrunn, S., & Bruning, R. (2023). Evaluating the effectiveness of writing interventions: Motivational outcomes. *Conceptualizing, Designing, Implementing, and Evaluating Writing Interventions,*. https://doi.org/10.1163/9789004546240_013

Authors and Affiliations

Naymé Salas[1] · Jofre Ventura[1] · Anna Llauradó[2]

✉ Naymé Salas
nayme.salas@uab.cat

Jofre Ventura
jofre.ventura@uab.cat

Anna Llauradó
allaura4@xtec.cat

[1] Departament de Didàctica de la Llengua i la Literatura, i de les Ciències Socials, Autonomous University of Barcelona, Cerdanyola del Vallès, Spain

[2] Departament d'Educació i Formació Professional, Barcelona, Spain