Effects of the STOP and LIST Strategy on the Writing Performance of Struggling Fourth Graders

Matthias Grunke[1], Kerstin Nobel[1] & Janine Bracht[1]

[1] Department of Special Education & Rehabilitation, University of Cologne, Germany

Correspondence: Matthias Grünke, Department of Special Education & Rehabilitation, University of Cologne, Klosterstr, 79b, 50931 Cologne, Germany. E-mail: matthias.gruenke@uni-koeln.de

Abstract

This study tested the effectiveness of a writing planning strategy (STOP & LIST) with four struggling students from fourth grade. A multiple-baseline design (AB) was used with baselines consisting of between five and seven daily probes. The duration of the intervention was between five and seven 25-minute sessions. A randomization procedure was implemented within the constraint that the baseline had to comprise at least five measurements and the treatment had to comprise at least three. The data were analyzed using visual inspection, different effect sizes, randomization tests, and piecewise regression analyses. Results revealed distinct improvements in both the length and quality of stories that the participants produced from baseline to the end of the training. This indicates that the ability of young struggling writers to plan narratives can be fostered considerably through rather simple means. Implications for practice and future research are discussed.

Keywords: writing problems, strategy instruction, single-case research, STOP and LIST

1. Introduction

"If speaking makes us human, writing makes us civilized" (Algeo, 2010, p. 6). The ability to create text using one's intellectual and linguistic resources is the primary foundation upon which record keeping, history, and art are grounded. It is a major means of expression, both for personal cognitive purposes and for communicating meaning with others. Without it, our society could not function (Fayol, Alamargot, & Berninger, 2012). Teaching children how to put their thoughts onto paper is indispensable for helping them to develop clarity of thought, to construct meaning, and to acquire reasoning skills (Graham, MacArthur, & Fitzgerald, 2013). In school, writing is the main instrument used to determine how much a student has learned and to assess his or her academic performance. This is true not only for language arts but for any subject that requires linguistic exposition (Mercer, Mercer, & Pullen, 2011; Saddler & Asaro-Saddler, 2010).

Even though the ability to transfer ideas into written form is crucial for the success of an individual in and out of school, as well as for the continuance of our society, according to Troia (2009), it is the most neglected of the basic skills taught during elementary or secondary education. Many teachers tend to stress reading, spelling, and arithmetic over text production. One reason is that they feel overwhelmed by the task of having to motivate their students to tackle such an extremely demanding activity and to instruct them on how to execute it (Gündoğmuş, 2018). Regrettably, a significant share of children and adolescents are unable to acquire ample writing skills without well-structured and frequent support (Graham & Harris, 2018).

Thus, it is no wonder that the prevalence rates of students not meeting basic standards in text production and qualifying for a so-called written language disability (WLD) are quite high. Children and adolescents with a WLD demonstrate significantly lower writing abilities than expected based on their chronological age, their measured intelligence, and the education that they have received so far (Nielsen et al., 2018). The incident rate of a WLD in students up to age 19 varies between 6.9% and 14.7% (Katusic, Colligan, Weaver, & Barbaresi, 2009).

To effectively aid children and adolescents to overcome their difficulties or to prevent them from significantly falling behind in their performance, it is essential to be aware of expedient starting points for an intervention. The most influential theories on writing (e.g., Bereiter & Scardamalia, 1987; Hayes, 1996; Hayes & Flower, 1980; Rijlaarsdam & van den Bergh, 1996) help to make the writing process transparent and palpable. In simple words, they subdivide the corresponding activities into three recursive phases: (a) planning (setting goals,

retrieving relevant information from long-term memory, generating ideas, and organizing ideas), (b) translating (putting ideas into visible language), and (c) revising (reading and editing what has been written so far). All of these models acknowledge that individuals plan, translate, and revise repeatedly. Thus, a writing activity is generally not a process in which an individual undergoes the various stages consecutively.

However, it is safe to say that generating, choosing, and structuring ideas occurs more often at the beginning of the endeavor than during the rest of the time. Research has shown that struggling writers tend to neglect planning activities in particular (e.g., Rodríguez, Grünke, González-Castro, García, & Álvarez-García, 2015). In fact, MacArthur and Graham (1987) found that they generally invest less than one minute in such prewriting efforts. Because those students do not take the time to brainstorm and organize their ideas, they basically jot down whatever comes to mind. As a result, they usually submit relatively short, incomplete, and incoherent texts (Englert & Raphael, 1988; MacArthur & Graham, 1987).

Due to how important writing is for students' overall success and the high number of them who struggle with composition writing, research on how to effectively support their endeavors is highly relevant. Fortunately, the basis of studies on the benefits of respective treatments is relatively strong by now. This is true not only for approaches aimed at facilitating translating and revising skills but also for treatments geared toward supporting the planning process. There are several meta-analyses on the effects of various writing interventions for struggling students (Cook & Bennett, 2014; Datchuk & Kubina, 2012; Gillespie & Graham, 2014; Gillespie Rouse & Sandoval, 2018; Rogers & Graham, 2008). Most of the studies included in these synopses were single-case analyses involving K–12 students with documented learning disabilities (LDs). The great majority of research papers utilized the well-known Self-Regulated Strategy Development (SRSD) model by Graham and Harris (1995), a six-stage approach of instruction: (a) activate background knowledge, (b) discuss the strategy, (c) model the strategy, (d) enable memorization of the strategy, (e) support the strategy, and (f) provide opportunities for independent strategy use. Based on explicit strategy instruction (see Reid, Linemann, & Hagaman, 2013), the SRSD model offers a framework for teachers to plan and deliver effective writing lessons. Overall, the findings on ways to assist with generating ideas and organizing them through the use of the SRSD model appear promising. The respective effect sizes focusing on the length and/or quality of narratives always ranged between medium and high (Grünke & Leonard Zabel, 2015; Sperling & Grünke, 2017).

Based on the findings of these meta-analyses, the so-called STOP and LIST strategy can be considered auspicious. However, it has received little attention in the scholarly literature so far. STOP and LIST is an acronym that stands for the following four stages: (a) Stop, (b) Think of Purpose, (c) List Ideas, and (d) Sequence Them (Graham & Harris, 2005). Based on the problem-solving theory by D'Zurillas and Goldfried (1971), STOP and LIST scaffolds the process of planning a writing product and teaches students how to generate and organize their ideas. The four stages are communicated by way of the SRSD model. STOP and LIST focuses on planning stories as the most basic text genre.

An extensive computer-supported search in the databases Educational Abstracts, ERIC, PsycINFO, Scopus, TOC Premier, and Web of Science yielded only three published studies about the effectiveness of STOP and LIST on the performance of students with writing problems: Graham, Harris, and Troia (1998); Grünke and Hatton (2017); and Troia and Graham (2002). In all these cases, the intervention was very beneficial in helping the participants to better plan and draft narratives and evoked sudden increases in achievement. The experiment by Graham et al. (1998) involved three fifth graders with LDs, the one by Grünke and Hatton (2017) included one boy with a specific WLD in sixth grade, and the one by Troia and Graham (2002) involved 20 fourth- and fifth-grade students. Supplying struggling learners with a road map of how to plan their texts is a provision that should not be held back for longer than necessary. As soon as children have acquired ample accuracy and fluency in basic transcription skills (spelling, handwriting, grammatical knowledge, etc.), they are equipped to learn how to produce simple stories from a first-person perspective. This usually happens during the late stages of their elementary school career (Berninger, Fuller, & Whitaker, 1996; Hacker, 2018; Kim & Schatschneider, 2017).

The purpose of the present study was to add to the scarce body of research on the effectiveness of the STOP and LIST strategy with struggling students in fourth grade. We deliberately chose at-risk children in an early phase of their writing development to precociously counteract severe problems that were likely to emerge if they were left to their own resources. It was expected that the participants would produce not only longer but also qualitatively better stories over the course of the intervention. In line with previous studies, we anticipated a sudden rather than a gradual increase in achievement.

2. Method

2.1 Participants and Setting

The participants were four 9-year-old fourth graders from a public inclusive elementary school in a major city in Northrhine-Westfalia: Adal, Berta, Channa, and Daniel (names changed for anonymity). A female graduate student of special education, who served as interventionist for this study, selected the children for participation based on recommendations from their main teacher. She had extensive experience working with children through various internships lasting several weeks in different schools. To be included in the experiment, the participants had to show severe deficits in writing composition. However, they had to be capable of composing at least simple sentences. In addition, they needed to demonstrate adequate handwriting and spelling skills. By resorting to the school records, the main teacher recommended the aforementioned children for participation.

To back up the proposals, the graduate student conducted the General German Language Test (Steinert, 2011), in which participants had to write a story in response to a prompt that consisted of a picture showing a man climbing up a ladder to a balcony and a woman watching him from a window in a neighboring house. There was no time limit for finishing the task. Results revealed that Adal, Berta, Channa, and Daniel were among the five students in the class who produced the shortest stories. Even though they had never been diagnosed with a disability and did not receive specialized instruction at the time of the study, they could clearly be considered to be at risk for academic failure in the area of written expression.

Adal is the son of Turkish migrant workers. Channa's parents were both born in Poland and migrated to Germany when she was a toddler. Berta and Daniel did not have immigrant backgrounds. Results from informal observations by the graduate student during two school visits in preparation for this research indicated that all four participants were easy to distract and difficult to motivate.

2.2 Design and Measures

We applied a single-case multiple baseline across participants design (AB) to investigate treatment effects before and during instruction (Kazdin, 2010). The whole experiment lasted only 12 consecutive school days. During this time, we collected 12 measurements from each participant. To increase the internal validity of the study (see Dugard, 2013), the start and close of the intervention were determined randomly for each case within the constraint that phase A had to consist of at least five and phase B of at least three probes. To detect autocorrelation in the data, it is often suggested to set the minimum number of baseline probes at five (e.g., Smith, 2012). Previous research gives rise to the hope that instruction in the STOP and LIST strategy does not have to take long to elicit notable improvements (Grünke & Hatton, 2017). Considering the aforementioned confinements, the treatment could have started any time between the sixth and tenth probe. A random drawing of all five possible options resulted in an arrangement whereby the training started on the day of the sixth measurement for Adal and Channa and on the day of the eighth measurement for Berta and Daniel.

Two dependent variables were applied to evaluate treatment effects: total words written (TWW) and a qualitative writing rubric (QWR). For baseline and intervention writing probes, participants were asked to produce a story in response to a randomly chosen picture from a pool of 20 DIN A4-size photographs. These depicted social scenes such as parties, conversations between people, or sports events (the pictures are available upon request). The students had as much lined DIN A4 notepaper at their disposal as they wanted. There were no time limits given to complete the assignment.

TWW was defined as the number of words written in a story (see Furey, Marcotte, Hintze, & Shackett, 2016). Titles were not included. However, incorrectly spelled, nonsense, or illegible words all counted toward TWW. To ensure accuracy, TWW was counted twice on two consecutive days by a female student assistant who was blind to the study. If the two counts differed (which happened in less than 5% of the runs), she counted a third time. A fourth count was never necessary.

The QWR was adapted by Glaser (2004) and constitutes a German variation of the Scale for Scoring the Inclusion and Quality of the Parts of a Story by Harris and Graham (1996). The assessment form comprised eight categories, each measurable from 0 to a maximum of 3 points. The first category was the protagonist of the composed text. Zero points were given if the student did not mention a main character at all. Up to 3 points were distributed if the student provided at least one detail of the protagonist in at least two following sentences. The next category was the setting of the text and ranged from not mentioning the setting to mentioning the setting with at least one detail in two following sentences. The same distribution of points was adapted to the aspect of time in which the text is set. The next category referred to the intention of the text's protagonist. Zero points were given if no intention was obvious, and up to 3 points were distributed if the student provided the reader

with a highly detailed and complex description of the protagonist's intention. The distribution of points for the plot were based on if the plot did not follow any logical order up to the plot following a logical order with at least five sentences written. The penultimate category was the ending of the text. Zero points were given if no ending was mentioned, which means no conflict resolution or dénouement was provided. Up to 3 points were distributed if a plot-orientated ending was described using direct speech, thoughts, or feelings of the protagonist. The last category involved the coherence of the written text. Zero points were given if there was no context and/or an incomplete structure so that parts of the plot were missing. Up to 3 points were distributed if the student provided the reader with a completely coherent text.

The range of total points that could be earned for a narrative ranged between 0 and 24. After receiving ample instructions on how to use the instrument from the first author, a female graduate student conducted the scorings. An experienced student assistant appraised a random selection of 20% of the texts independently. We calculated the interrater agreement for each of these writing products by dividing the smaller number of points by the larger one and multiplying by 100. Average interrater reliability reached a remarkably high level of 94.95%. This degree of agreement gave us the confidence to go by the graduate students' appraisals and view them as dependable.

2.3 Procedures

At the beginning of the second, third, fourth, and fifth period of each day of the experiment (each lasting 45 minutes), the previously mentioned interventionist took one of the four participants into a resource room of the school, while the rest of the children remained in class with their main teacher. Even though our four subjects were not the only students present, it was always possible to find a quiet place away from children not receiving treatment. The order in which Adal, Berta, Channa, and Daniel were asked to go with the interventionist varied daily. During baseline conditions, the students were given coloring books to keep them occupied for 20 minutes. In the remaining 25 minutes, the children were asked to produce a story based on a randomly chosen picture, as outlined in the section on design and measures.

During intervention, the participants still had to write a text at the end of a given period. However, prior to this, instead of coloring books, they were instructed in the use of the STOP and LIST strategy. Neither in phase A nor in phase B did Adal, Berta, Channa, or Daniel need more than 15 minutes to finish their stories. Whenever they indicated that they were done, the interventionist took them back to their classroom. Contrary to producing texts during baseline, all participants started taking notes in accordance with the STOP and LIST strategy before writing a narrative as soon as the treatment had commenced.

In accordance with the SRSD model, Lesson 1 started with guiding the children to recall their basic knowledge of strategies and aspects of focus when composing a text. This led to an initial reflection of the students' writing process and enabled them to recall certain rules they had learned so far (e.g., "Be organized," "Take good notes", "Ask for help if you get stuck"). Secondly, the interventionist promised that the participants would be able to write better stories if they assiduously applied the planning strategy she was about to introduce. Subsequently, she modelled the process of working through the four steps of STOP and LIST, using a writing prompt in the form a DIN A4-size printed picture taken from https://www.teachstarter.com/widget/visual-writing-prompts/. While going through the various phases, the interventionist continuously referred to a DIN A3 poster that she placed on the table and that contained the instructions of the strategy (Stop, Think of Purpose, List Ideas, Sequence Them). In each incidence, the interventionist produced six ideas on what could have happened in a story related to the photograph, jotted them down, and organized them in a chronological sequence. The students were asked to be aware that they now had a useful outline for producing a more or less detailed and captivating narrative. In closing, the interventionist told a story based on her notes while trying to develop a comprehensible plot with a clear beginning, middle, and end. She advised the children that they should always make a draft of what they want to write about before actually putting down a text. At the end of this and all lessons in which the poster was used, it was taken away before performance was measured.

Lesson 2 started with the interventionist recapitulating the strategy steps while pointing to the poster that was again placed on the table. Afterward, the interventionist turned the poster over and asked the participants to recall the four stages by heart. If a child got stuck, the interventionist offered assistance. As soon as a student managed to repeat the steps without a mistake, another writing prompt in the form of a photograph was chosen to create another story outline. This time, the interventionist provided guided practice while she and the students went through the process together. As soon as they finished, they repeated the procedure with yet another writing prompt.

In Lesson 3, the participants were asked to produce a story outline by themselves. The interventionist placed the

poster on the edge of the table to serve as a memory aid. Insofar as the participants were capable of applying the strategy independently, the interventionist withheld her help. As soon as a child did not know what to do next, she intervened. By the end of Lesson 3, all students were able to apply the strategy steps, end up with four to eight sequenced notes, and tell a story based on their memos.

During the remaining lessons, the participants kept practicing STOP and LIST with the interventionist supporting them as far as it seemed expedient. The poster was not used from this time onward. However, the interventionist reminded the children of the four strategy steps any time they did not seem to know what to do next.

2.4 Treatment Fidelity

The first author extensively instructed the interventionist during three 1-hour lessons prior to the beginning of the treatment using exemplary video images depicting the first author applying the strategy with different students. In addition, the graduate student was given a detailed script to follow and a checklist that contained every important step of the procedures (both can be obtained from the authors upon request). She was asked to record her actions using this list to ensure that lessons were delivered in accordance with the plan. During the intervention, she stayed in daily contact with the first author via e-mail or phone to discuss the progress of the treatment and to make sure that it was implemented as intended.

3. Results

Descriptive data for the TWW and QRW results are displayed in Tables 1 and 2. Both measures are highly correlated across cases ($r_{Spearman's Rank}$ = .87; p < .001; one-tailed).

Table 1. Total words written for each participant

		Baseline	Intervention
Adal	*N* (Probes)	5	7
	Raw Scores	56; 46; 43; 44; 50	79; 124; 121; 129; 136; 189; 144
	M	47.80	131.71
	SD	5.31	32.69
	Range	43–56	79–189
Berta	*N* (Probes)	7	5
	Raw Scores	41; 40; 30; 41; 36; 39; 31	149; 189; 184; 166; 190
	M	36.86	175.60
	SD	4.67	17.73
	Range	30–41	149–190
Channa	*N* (Probes)	5	7
	Raw Scores	40; 45; 46; 34; 39	82; 92; 98; 107; 110; 104; 112
	M	40.80	100.71
	SD	4.87	10.78
	Range	34–46	82–112
Daniel	*N* (Probes)	7	5
	Raw Scores	32; 36; 39; 33; 27; 28; 40	95; 90; 81; 86; 96
	M	33.57	89.60
	SD	5.06	6.27
	Range	27–40	81–96

Table 2. Qualitative writing rubric for each participant

		Baseline	Intervention
Adal	*N* (Probes)	5	7
	Raw Scores	8; 6; 6; 5; 8	9; 14; 9; 13; 19; 15; 12
	M	6.60	13.00
	SD	1.34	3.51
	Range	5–8	9–19
Berta	*N* (Probes)	7	5
	Raw Scores	5; 3; 5; 4; 7; 7; 4	15; 14; 14; 18; 15
	M	5.00	15.20
	SD	1.53	1.64
	Range	3–7	14–18
Channa	*N* (Probes)	5	7
	Raw Scores	5; 8, 6; 2; 4	9; 13; 14; 9; 9; 16; 13
	M	5.00	11.85
	SD	2.24	2.85
	Range	2–8	9–16

Daniel	N (Probes)	7	5
	Raw Scores	6; 4; 7; 3; 6; 5; 8	12; 13; 14; 12; 14
	M	5.57	13.00
	SD	1.72	1.00
	Range	3–8	12–14

A visual display of the data is provided in Figures 1 and 2. All graphs were produced using the SCAN package for R by Wilbert (2018). A visual inspection of the data reveals fairly stable baselines for all four participants and both dependent measures. It also suggests a sudden increase in performance upon the onset of the intervention.
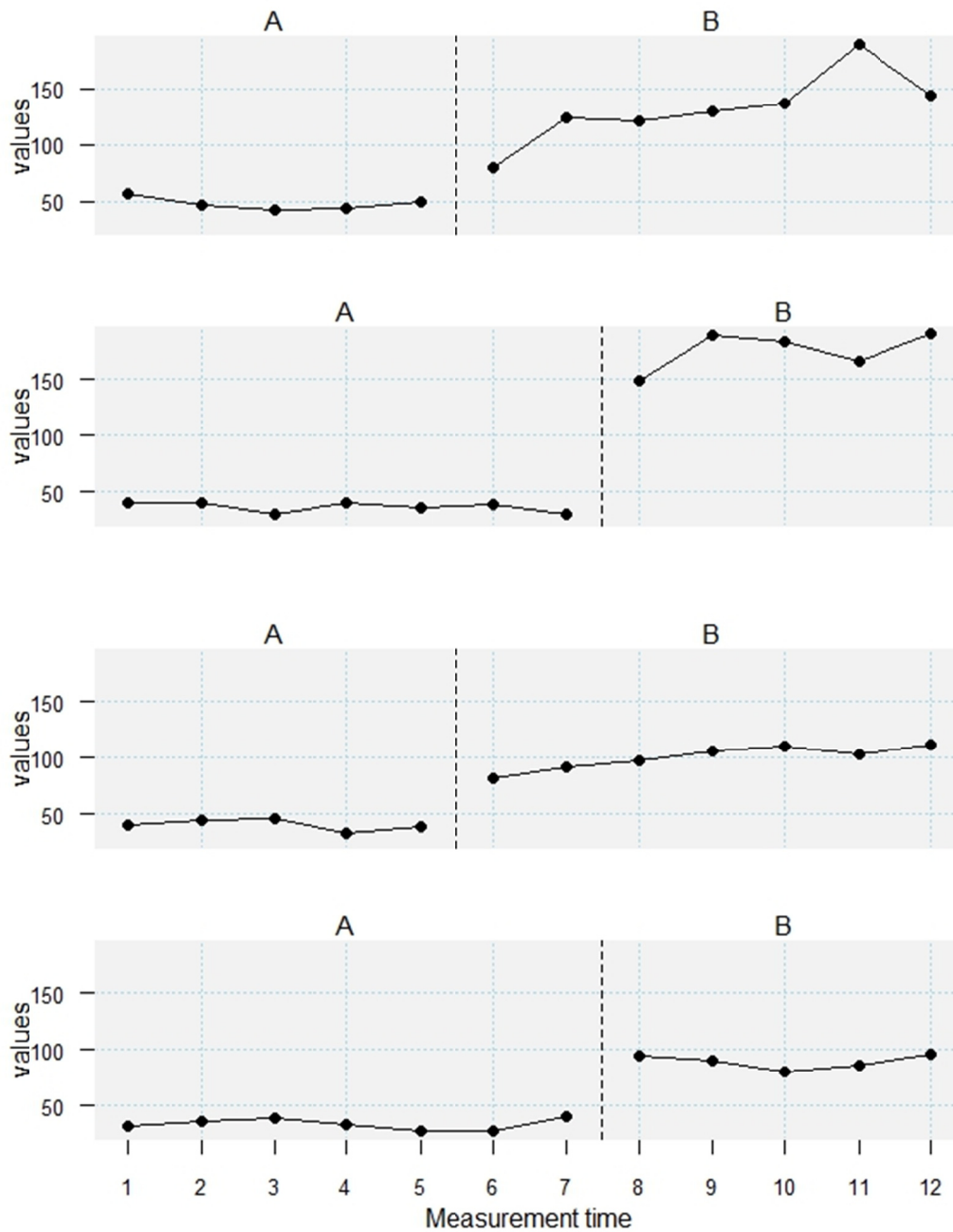

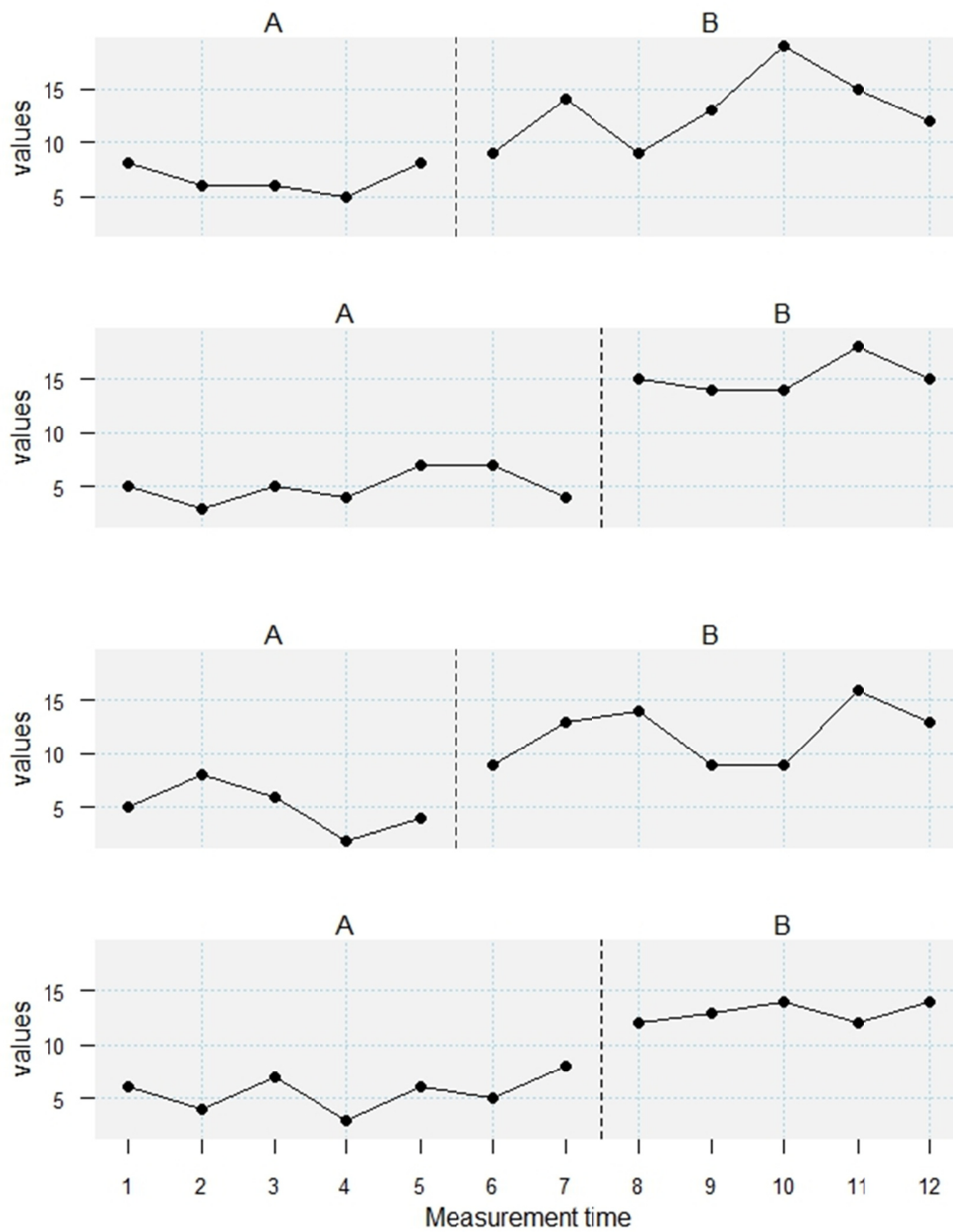Figure 1. Total words written for Adal, Berta, Channa, and Daniel

Figure 2. Qualitative writing rubric for Adal, Berta, Channa, and Daniel

Table 3 presents the results for some of the most common effect size measures used in single-case research: percentage of non-overlapping data (PND; Scruggs, Mastropieri, & Casto, 1987), percentage of all non-overlapping data (PAND; Parker, Hagan-Burke, & Vannest, 2007), improved rate difference (IRD; Parker, Vannest & Brown, 2009), mean baseline difference (MBD; Cambell, 2003), Glass's Δ (Glass, McGaw & Smith, 1981), non-overlap of all pairs (NAP; Parker & Vannest, 2009), and Tau-U (Parker, Vannest, Davis, & Sauber, 2010). The indices were calculated using the SCDA package, a plug-in for R-Commander, developed by Bulté (2013; for MBD and Glass's Δ), and the SCAN package by Wilbert (2018; for all non-overlap effect sizes).

Because there was no overlap between the values in phases A and B for any of the four subjects, the indices for PND, PAND, IRD, NAP, and Tau-U each reached the highest score possible. For NAP and Tau-U, the SCAN

package provides p-values and confidence intervals. As indicated in Table 3, all respective data analyses yielded statistically highly significant differences between phases A and B. The two effect size measures not quantifying overlap (MBD and Glass's Δ) offer a more differentiated picture. MBD is generated by subtracting the mean of the baseline points from the mean of the treatment points, then dividing by the mean of the baseline points and multiplying by 100. There is no commonly accepted convention for classifying different MBD values as small, moderate, or high. However, considering that an MBD index represents the percentage of performance increase, data between 97 and 376 can certainly be viewed as remarkable. Aside from the statement that all children showed notable increases, it can be asserted that Berta profited the most. In all cases, the improvements in TWW were greater than in the qualitative appraisals.

Table 3. Effect sizes for total words written (TWW) and qualitative writing rubric (QWR)

	PND	PAND	IRD	MBD	Glass's Δ	NAP	Tau-U
TWW							
Adal	100	100	1.00	176	15.80	1.00***	1.00***
Berta	100	100	1.00	376	29.71	1.00***	1.00***
Channa	100	100	1.00	147	12.31	1.00***	1.00***
Daniel	100	100	1.00	167	11.07	1.00***	1.00***
QWR							
Adal	100	100	1.00	97	4.77	1.00***	1.00***
Berta	100	100	1.00	204	6.68	1.00***	1.00***
Channa	100	100	1.00	137	3.07	1.00***	1.00***
Daniel	100	100	1.00	133	4.32	1.00***	1.00***

Note. PND=percentage of non-overlapping data; PAND=percentage of all non-overlapping data; IRD=improved rate difference; MBD=mean baseline difference; NAP=non-overlap of all pairs.

Glass's Δ is considered an appropriate effect size measurement if baseline data are not excessively variable and do not present trends. It expresses the difference between intervention and baseline means in units of the standard deviation of the baseline scores. The convention for categorizing Cohen's d (1988) indices into small (0.20), medium (0.5), and large (0.80) does not apply to Glass's Δ. However, scores between 11.07 and 29.71 for quantitative measures (TWW) can be considered tremendously high. Indices between 3.07 and 6.68 for the qualitative appraisals (writing rubric score) are not enormous, but still remarkable. According to the Glass's Δ scores for TWW, Berta benefitted the most (29.71) and Daniel the least (11.07). Based on the assessments done with the writing rubric, Berta profited the most from the training (6.68), whereas Channa (3.07) showed the smallest gains.

Furthermore, we applied a randomization test suitable for multiple-baseline designs (AB; Edgington & Onghena, 2007), again using the SCAN package for R by Wilbert (2018). Randomization tests are not yet widely spread among researchers engaged in single-case research. They are model-free, computationally intensive methods for hypothesis testing that generate many replicates of an actual dataset—typically called pseudo-samples—and use the pseudo-samples to estimate a score distribution (Grünke, Boon, & Burke, 2015; La Fond & Neville, 2010). Randomization tests offer the great advantage of enabling analysis of an overall effect for all participants and providing exact p-values for generalizing the results (Dugard, 2013). For TWW, data analysis yielded a highly significant effect (ΔM_{AB}=123.04–39.00=84.04; p<.01); for QWR, it generated at least respectively significant evidence (ΔM_{AB}=13.13–5.50=7.63; p<.05).

Finally, we conducted piecewise regression analyses to explore the relevance of different effects (trend, level, and slope) for the results (Moeyaert, Ugille, Ferron, Beretvas, & Van Den Noortgate, 2014). Again, we utilized the SCAN package by Wilbert (2018) to execute the calculations. The outcomes of the analyses are provided in Tables 4 and 5.

Table 4. Piecewise regression model for total words written

	B	SE	t	p	R²
		Adal			
Intercept	52.00	18.08	2.88	.02*	
Trend	-1.40	5.45	-0.26	.80	.00
Level	38.14	19.76	1.93	.09	.05
Slope	13.54	6.35	2.13	.07	.05
		Berta			
Intercept	40.57	9.52	4.26	.00***	
Trend	-0.93	2.13	-0.43	.67	.00
Level	123.83	14.09	8.79	.00**	.17
Slope	6.83	4.15	1.65	.14	.01
		Channa			
Intercept	44.70	5.35	8.35	.00***	
Trend	-1.30	1.61	-0.81	.44	.00
Level	44.51	5.85	7.61	.00**	.13
Slope	5.80	1.88	3.09	.02*	.02
		Daniel			
Intercept	34.14	5.26	6.49	.00***	
Trend	-0.14	1.18	-0.12	.91	.00
Level	57.06	7.79	7.33	.00***	.22
Slope	-0.06	2.29	-0.03	.98	.00

Note. * significant at the .05 level. ** significant at the .01 level. *** significant at the .001 level.

Table 5. Piecewise regression model for qualitative writing rubric

	B	SE	t	p	R²
		Adal			
Intercept	6.90	3.00	2.30	.05*	
Trend	-0.10	0.90	-0.11	.92	.00
Level	3.60	3.28	1.10	.30	.05
Slope	0.85	1.05	0.81	.44	.03
		Berta			
Intercept	4.00	1.38	2.89	.02*	
Trend	0.25	0.31	0.81	.44	.01
Level	8.25	2.05	4.03	.00***	.13
Slope	0.15	0.60	0.25	.81	.00
		Channa			
Intercept	7.40	2.79	2.66	.03*	
Trend	-0.80	0.84	-0.95	.37	.03
Level	6.60	3.05	2.17	.06	.16
Slope	1.26	0.98	1.29	.02*	.06
		Daniel			
Intercept	4.57	1.31	3.50	.01**	
Trend	0.25	0.29	0.86	.42	.01
Level	5.78	1.93	2.99	.02*	.12
Slope	0.05	0.57	0.09	.93	.00

Note. * significant at the .05 level. ** significant at the .01 level. *** significant at the .001 level.

First, our initial impression of a stable baseline was confirmed—there was no significant trend for any of the students in phase A. Moreover, as shown in Table 4, for all but one participant, there was a highly significant level effect relating to phase differences in TWW, with R^2 between 13 and 22%. In addition, Channa showed a noteworthy slope effect. However, no significant results were identified for Adal. With regard to improvements in writing quality (see Table 5), only Berta and Daniel demonstrated a meaningful increase in level.

Even though the piecewise regression analyses did not support all hypotheses proposed in this study (the level effect was only significant in five out of eight cases), the results need to be seen as a whole. And taking all modes of data analysis into account (visual inspection, effect size estimates, randomization tests, and piecewise regression modelling), the picture becomes relatively clear: The intervention certainly had a distinct impact on the scores of the participants.

4. Discussion

4.1 Basic Findings

This study aimed at shedding lighter on the effectiveness of a promising strategy to help struggling writers plan better stories (STOP & LIST). The results of our single-case analysis appear very encouraging. Visual inspection indicates a remarkable quantitative and qualitative improvement from baseline to intervention. In addition, all effect size measures point to notably higher achievements as soon as the participants were taught the strategy. Finally, inferential statistics (randomization test and piecewise regression analyses) substantiate these appraisals of the treatment benefits. The level effects in particular were striking.

Thus, our findings tie in with the ones produced by other studies focusing on STOP and LIST (Graham, Harris, & Troia 1998; Grünke & Hatton, 2017; Troia & Graham, 2002). The results of this experiment fortify the assumption that the lack of planning is a crucial barrier for struggling writers as they try to produce stories of an acceptable length and of sufficient quality. As soon as they receive some substantiated instruction on how to generate and organize ideas, their performance increases considerably. The insights from this research agree with all major writing theories that recognize planning as an essential part of the process (see MacArthur, Graham, & Fitzgerald, 2006), as well as with the findings from existing meta-analyses on the subject (Cook & Bennett, 2014; Datchuk & Kubina, 2012; Gillespie & Graham, 2014; Gillespie Rouse & Sandoval, 2018; Rogers & Graham, 2008).

4.2 Limitations

Like all empirical research, this study is subject to certain limitations. First, the results of an intervention with only four students do not allow for far-reaching conclusions about how to successfully support every struggling writer. Additionally, our participants were all the same age and had a similar educational background. Thus, our findings can only refer to students of this particular sub-segment. It has to be taken into account that children struggling with text composition are a heterogeneous group. Not every one of them necessarily benefits from one certain intervention. It is yet to be determined whether STOP and LIST has the potential to help a wider variety of learners than just students of a specific age group with moderate deficits. None of the participants of this study was ever diagnosed with a WLD. In fact, the purpose of the treatment applied was to prevent the problems from expanding and intensifying. Even though previous studies have focused on students with LDs or WLDs (Graham et al., 1998; Grünke & Hatton, 2017), the research base is anything but broad enough to call STOP and LIST evidence-based for a variety of subgroups of students with writing problems. It would be beneficial to apply this technique to a larger group of children and adolescents with different characteristics to identify responders and nonresponders.

Another limitation concerns the design of our study. We applied a multiple-baseline plan (AB) without collecting any follow-up data. Therefore, it is not possible to draw any inferences about the long-term effects of the intervention on the writing performance of the participants. However, considering that teaching our students how to plan their stories had an immediate impact on their performance gives rise to the hope that teaching a learner how to use the tool will help him or her not only during a treatment phase but even beyond.

A final limitation pertains to the way we captured writing performance. Assessing this skill is generally more complicated than detecting reading fluency or spelling. This is especially true when trying to measure the quality of a text, which always results in relatively subjectively colored data. Using writing rubrics is the most conventional mode to monitor the degree of excellence in a text over time. However, the sheer quantity of these scoring guides is dizzying. Almost every single-subject experiment in the previously cited meta-analyses used its own self-constructed tool for determining writing quality. Thus, there is obviously no silver bullet among the rubrics. Even though this was dissatisfactory, we had to choose one particular instrument from among a variety of alternatives to measure writing quality, knowing full well that any decision in this respect would make us prone to criticism.

4.3 Implications for Practice

Writing skills have great relevance for the school careers of children and adolescents as well as for their social participation. Luckily, research suggests that students struggling with this skill can be supported effectively. This study confirms that STOP and LIST is a very promising tool to improve their writing. It is easy to implement in educational settings because using STOP and LIST is highly intuitive, and teaching it requires no special equipment. Even peer-tutoring concepts, that compensate for a possible lack of ideas in students who struggle with composition writing, can disburden the teacher and leave capacity for support where needed elsewhere. In this way, STOP and LIST can be applied in very heterogeneous classrooms. Because the main part of the STOP

and LIST strategy is to find and sequence ideas for the subsequent writing task, one option could be to divide the class into cooperative learning settings, such as groups or partners, to enhance the quantity and quality of ideas the students generate. To meet the increasing influences of technology-enhanced learning, the authors are in the middle of developing a digital version of the STOP and LIST strategy.

4.4 Conclusion

Further research on this promising strategy could enhance the insights into the improvement of writing competence. Future studies should focus on replicating this experiment while addressing the limitations discussed above. The findings strongly support the hypothesis that STOP and LIST has positive effects on the quantitative and qualitative writing performance of struggling students.

References

Algeo, J. (2010). *The origins and development of the English language*. Boston, MA: Wadsworth Cengage Learning.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum.

Berninger, V. W., Fuller, F., & Whitaker, D. (1996). A process model of writing development across the life span. *Educational Psychology Review*, *8*, 193–218. https://doi.org/10.1007/BF01464073

Bulté, I. (2013). *Being grateful for small mercies: The analysis of single-case data*. Unpublished doctoral dissertation. KU Leuven, Belgium.

Campbell, J. M. (2003). Efficacy of behavioral interventions for reducing problem behavior in persons with autism: A quantitative synthesis of single-subject research. *Research in Developmental Disabilities*, *24*, 120–138. https://doi.org/10.1016/S0891-4222(03)00014-3

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Cook, K. B., & Bennett, K. E. (2014). Writing interventions for high school students with disabilities: A review of single-case design studies. *Remedial and Special Education*, *35*, 344–355. https://doi.org/10.1177/0741932514523140

Datchuk, S. M., & Kubina, R. M. (2012). A review of teaching sentence-level writing skills to students with writing difficulties and learning disabilities. *Remedial and Special Education*, *34*, 180–192. https://doi.org/10.1177/0741932512448254

Dugard, P. (2013). Randomization tests: Are they what you need? *Insights into Learning Disabilities*, *10*, 87–93.

D'Zurilla, T. J., & Goldfried, M. R. (1971). Problem solving and behavior modification. *Journal of Abnormal Psychology*, *78*, 107–126. https://doi.org/10.1037/h0031360

Edgington, E. S., & Onghena, P. (2007). *Randomization tests*. Boca Raton, FL: Chapman & Hall.

Englert, C. S., & Raphael, T. E. (1988). Constructing well-formed prose: Process, structure, and metacognitive knowledge. *Exceptional Children*, *54*, 513–520. https://doi.org/10.1177/001440298805400604

Fayol, M., Alamargot, D., & Berninger, V. W. (2012). *Translation of thought to written text while composing*. New York, NY: Taylor & Francis. https://doi.org/10.4324/9780203141434

Furey, W. M., Marcotte, A. M., Hintze, J., & Shackett, C. (2016). Concurrent validity and classification accuracy of curriculum-based measurement for written expression. *School Psychology Quarterly*, *31*, 369–382. https://doi.org/10.1037/spq0000138

Gillespie, A., & Graham, S. (2014). A meta-analysis of writing interventions for students with learning disabilities. *Exceptional Children*, *80*, 454–473. https://doi.org/10.1177/0014402914527238

Gillespie Rouse, A., & Sandoval, A. (2018). Writing interventions for students with learning disabilities: Characteristics of recent research. *Learning Disabilities: A Multidisciplinary Journal*, *23*, 1–17.

Glaser, C. (2004). *Förderung der Schreibkompetenz bei Grundschülern [Fostering writing competence in elementary school students]*. Münster, Germany: Waxman.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. London, UK; Sage.

Graham, S., & Harris, K. R. (1989). Improving learning disabled students' skills at composing essays: Self-instructional strategy training. *Exceptional Children*, *56*, 201–214. https://doi.org/10.1177/001440298905600305

Graham, S., & Harris, K. R. (2005). *Writing better. Effective strategies for teaching students with learning difficulties*. Baltimore: Paul H. Brookes.

Graham, S., & Harris, K. R. (2018). Evidence-based writing practices: A meta-analysis of existing meta-analyses. In R. Fidalgo, K. R. Harris, & M. Braaksma (Eds.), *Design principles for teaching effective writing: Theoretical and empirical grounded principles* (pp. 33–37). Leiden, Netherlands: Brill Academic Publishers.

Graham, S., Harris, K. R., & Troia, G. A. (1998). Writing and self-regulation: Cases from the self-regulated strategy development model. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practice* (pp. 20–41). New York, NY: Guilford.

Graham, S., MacArthur, C. A., & Fitzgerald, J. (2013). *Best practices in writing instruction*. New York, NY: Guilford.

Grünke, M., Boon, R. T., & Burke, M. (2015). Use of the randomization test in single-case research. *International Journal for Research in Learning Disabilities*, *2*, 44–64.

Grünke, M., & Hatton, H. (2017). Effects of the STOP and LIST strategy on the writing performance of a sixth grader with learning disabilities. *Insights into Learning Disabilities*, *14*, 155–165.

Grünke, M., & Leonard Zabel, A. M. (2015). How to support struggling writers. *International Journal of Special Education*, *30*, 137–150.

Gündoğmuş, H. D. (2018). Self-efficacy of teacher candidates for teaching first reading and writing. *Educational Research and Reviews*, *13*, 224–229. https://doi.org/10.5897/ERR2018.3486

Hacker, D. J. (2018). A metacognitive model of writing: An update from a developmental perspective. *Educational Psychologist*, *53*, 220–237. https://doi.org/10.1080/00461520.2018.1480373

Hartmann, D. P., & Pelzel, K. E. (2005). Design, measurement, and analysis in developmental research. In M. H. Bornstein & M. E. Lamb (Eds.), *Developmental psychology* (pp. 103–184). Mahway, NJ: Erlbaum.

Harris, K., & Graham, S. (1996). *Making the writing process work: Strategies for composition and self-regulation*. Cambridge, MA: Brookline.

Hayes, J. (1996). A new framework for understanding cognition and affect in writing. In M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1–27). Mahwah, NJ: Erlbaum.

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

Katusic, S. K., Colligan, R. C., Weaver, A. L., & Barbaresi, W. J. (2009). The forgotten learning disability: Epidemiology of written-language disorder in a population-based birth cohort (1976–1982). *Pediatrics*, *123*, 1306–1313. https://doi.org/10.1542/peds.2008-2098

Kazdin, A. E. (2010). *Single-case research designs: Methods for clinical and applied settings*. Oxford, UK: Oxford University Press.

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

La Fond, T., & Neville, J. (2010). *Randomization tests for distinguishing social influence and homophily effects* (pp. 601–610). Proceedings of the 19th International Conference on World Wide Web. WWW '10. New York, USA: ACM. https://doi.org/10.1145/1772690.1772752

MacArthur, C. A., & Graham, S. (1987). Learning disabled students' composing under three methods of text production: Handwriting, word processing, and dictation. *Journal of Special Education*, *21*, 22–42. https://doi.org/10.1177/002246698702100304

MacArthur, C. A., Graham, S., & Fitzgerald, J. (2006). *Handbook of writing research*. New York, NY: Guilford.

Mercer, C. D., Mercer, A. R., & Pullen, P. C. (2011). *Teaching students with learning problems*. Boston, MA: Pearson.

Moeyaert, M., Ugille, M., Ferron, J. M., Beretvas, S. N., & Van Den Noortgate, W. (2016). The misspecification of the covariance structures in multilevel models for single-case data. *Journal of Experimental Education*, *84*, 473–509. https://doi.org/10.1080/00220973.2015.1065216

Nielsen, K., Andria-Habermann, K., Richards, T., Abbott, R., Mickail, T., & Berninger, V. (2018). Emotional and behavioral correlates of persisting specific learning disabilities in written language during middle childhood and early adolescence. *Journal of Psychoeducational Assessment*, *36*, 651–669. https://doi.org/10.1177/0734282917698056

Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percent of all non-overlapping data (PAND): An alternative to PND. *The Journal of Special Education*, *40*, 194–204. https://doi.org/10.1177/00224669070400040101

Parker, R. I., & Vannest, K. J. (2009). An improved effect size for single case research: Non-overlap of all pairs (NAP). *Behavior Therapy*, *40*, 357–367. https://doi.org/10.1016/j.beth.2008.10.006

Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single case research. *Exceptional Children*, *75*, 135–150. https://doi.org/10.1177/001440290907500201

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2010). Combining non-overlap and trend for single case research: Tau-U. *Behavior Therapy*, *42*, 284–299. https://doi.org/10.1016/j.beth.2010.08.006

Reid, R., Lieneman, T. O., & Hagaman, J. L. (2013). *Strategy instruction for students with learning disabilities*. New York, NY: Guilford.

Rijlaarsdam, G., & van den Bergh, H. (1996). The dynamics of composing—An agenda for research into an interactive compensatory model of writing: Many questions, some answers. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 107–125). Hillsdale, NJ, US: Lawrence Erlbaum Associates.

Rodriguez, C., Grünke, M., González-Castro, P., Garcia, T., & Alvarez-Garcia, D. (2015): How do students with attention-deficit/hyperactivity disorders and writing learning disabilities differ from their non-labeled peers in the ability to compose texts? *Learning Disabilities: A Contemporary Journal*, *13*, 157–175.

Rogers, L. A., & Graham, S. (2008). A meta-analysis of single-subject design writing intervention research. *Journal of Educational Psychology*, *100*, 879–906. https://doi.org/10.1037/0022-0663.100.4.879

Saddler, B., & Asaro-Saddler, K. (2010). Written expression. In R. T. Boon, & V. G. Spencer (Eds.), *Best practices for the inclusive classroom* (pp. 187–204). Waco, TX: Prufrock.

Scruggs, T., Mastropieri, M., & Casto, G. (1987). The quantitative synthesis of single subject research. *Remedial and Special Education*, *8*, 24–33. https://doi.org/10.1177/074193258700800206

Smith, J. D. (2012). Single-case experimental designs: A systematic review of published research and current standards. *Psychological Methods*, *17*, 1–70. https://doi.org/10.1037/a0029312

Steinert, J. (2011). *General German Language Test (GGLT)*. Göttingen, Germany: Hogrefe.

Sperling, M., & Grünke, M. (Eds.). (2017). Preparing students with learning disabilities for writing tasks [Special issue]. *Insights into Learning Disabilities*, *14*.

Troia, G. (2009). Introduction. In G. Troia (Ed.), *Writing instruction and assessment for struggling writers: From theory to evidenced-based practices* (pp. 1–11). New York, NY: Guilford.

Troia, G. A., & Graham, S. (2002). The effectiveness of a highly explicit, teacher-directed strategy instruction routine: Changing the writing performance of students with learning disabilities. *Journal of Learning Disabilities*, *35*, 290–305. https://doi.org/10.1177/00222194020350040101

Wilbert, J. (2018). *Package "SCAN": Single-case data analyses for single and multiple AB designs.* Retrieved from https://www.uni-potsdam.de/fileadmin01/projects/inklusion/scan/scan.pdf

Copyrights