The Effects of a Multicomponent Strategy Instruction on the Argumentative Writing Performance of Low-Achieving Secondary Students

Susanne Hisgen*
Anne Barwasser
Thorben Wellmann
Matthias Grünke
University of Cologne

The ability to compose argumentative essays of sufficient quality is tantamount to the aptitude of thinking independently and critically, which is vital to playing an active part in one's society. Thus, the aim of teaching students this skill is incorporated into virtually every secondary school curriculum in almost all democratic countries around the globe. Unfortunately, adolescents with learning disabilities and other academic challenges find it especially difficult to write texts that present arguments about both sides of a controversial issue and weigh up the evidence before taking a stance. In this study, we investigated the effectiveness of self-regulated strategy development for an approach called STOP & DARE with 77 underperforming secondary students. Because academically challenged learners find it particularly difficult to engage in such an arduous task as acquiring argumentative essay skills, we complemented our intervention concept with some motivational techniques. We conducted our study within a randomized control group design: While half of the sample received the STOP & DARE training, the other half took part in a mindfulness course. Results showed a clear increase in performance among the students who attended the writing intervention sessions, whereas the control group's skill level rose only negligibly. These effects were still evident after 4 weeks. This suggests that the ability of struggling secondary students to compose admissible argumentative essays can be significantly improved even with relatively little means. We discuss these findings and their implications in relation to the possibilities of implementing STOP & DARE under everyday conditions in school.

Keywords: **Argumentative Writing, Self-Regulated Strategy Development, Learning Disabilities, STOP & DARE, Secondary School Students**

Introduction

The ability to write is an immensely important cultural skill beyond the scope of school (Graham, 2007). Proficiency in putting thoughts on paper is a central competence and a main requirement for social participation (Graham & Perin, 2007a; O'Brien et al., 2007). As one key predictor of academic success, writing has

---

*Please send correspondence to: Susanne Hisgen, Department of Special Education & Rehabilitation, University of Cologne, Klosterstr. 79b, Cologne, Northrhine-Westfalia, 50931, Germany, Phone: +49 (0)221-470 2124, Email: shisgen@uni-koeln.de.

to be taught extensively. Schools have an obligation to provide opportunities where students learn how to produce adequate texts (Harris et al., 2013).

Being able to express one's ideas in written form is a capability in high demand, and the acquisition of appropriate skills is a long-term successive process. According to the developmental model by Mayer (2007), children who are on track to develop age-appropriate text-production skills begin their first attempts at writing between the ages of 5 and 7. They pass through different stages of complexity until achieving an adequate competency level. Hayes and Flower (1980) underlined the complexity of this iterative process by stating that successful writers must perform three recursive steps during the act of composing: planning, translating, and revising. All of them are controlled by a monitor that coordinates these processes with respect to suitable time periods for each phase. One must attend to all of these features to arrive at a text of sufficient quality (Fayol et al., 2012).

Among all the different genres one can write within, argumentative essays are certainly among the most ambitious. They are also known as opinion or position papers. This writing genre requires an individual to form a personal standpoint on a topic after weighing up different arguments and coming up with a conclusion supported by credible evidence (Mason & Shriner, 2008). Composing an argumentative essay could be viewed as very similar to the process of problem-solving (Bereiter & Scardamalia, 1987). This too demands the use of goal-directed and self-regulated actions. Thus, teaching students to produce coherent and conclusive position papers is equivalent to teaching them to think critically and to generate potential strategies for problem resolution (Graham et al., 2013).

It would be desirable if all students met the standards for writing acceptable argumentative essays. However, this places a heavy strain on one's working memory. Unfortunately, a distressing number of adolescents in secondary schools are not even capable of producing simpler text forms such as narratives (Katusic et al., 2009). Within the group of struggling writers, those with learning disabilities (LDs) are certainly the largest (Troia, 2009). Studies have shown that students with LDs are barely able to plan or revise their texts and experience trouble expressing their ideas (Santangelo, 2014).

Luckily, since the 1980s, an increased focus has been placed on providing teachers with the means to effectively help struggling children and youth develop better writing abilities, including better opinion essay-production skills (i.e., Graham & Harris, 2005; Graham & Perin, 2007b). In evaluating the different options, strategy instruction has been shown to be especially effective for academically challenged learners (Graham et al., 2012; Graham & Perin, 2007b; Graham et al., 2013; Rogers & Graham, 2008). This category of intervention approaches refers to a systematic and explicit teaching of steps or actions that must be taken by individuals to improve their planning, translating, or revising skills (Graham, 2006). In their comprehensive meta-analysis, Graham et al. (2015) documented that strategy instruction is one of the most effective techniques to foster writing abilities in both students without LDs ($d = .93$) and students with LDs ($d = 1.00$).

The common framework for teaching techniques or devices that helps learners acquire text-production skills is the self-regulated strategy development (SRSD) model by Graham and Harris (2000). It includes procedures for goal-setting, self-

monitoring, self-instruction, and self-reinforcement. SRSD instruction consists of six phases: (1) develop background knowledge; (2) discuss strategy; (3) model the strategy; (4) memorize the strategy; (5) support the strategy; and (6) allow the learners to perform the strategy independently (Ennis & Jolivette, 2013; Harris et al., 2008). The meta-analysis by Gillespie and Graham (2014) showed that teaching writing within the framework of SRSD has a remarkably high effect ($d = 1.33$).

Among all the intervention approaches that have been developed and evaluated, those that aim to improve the vital skill of argumentative essay writing are astonishingly scarce. One rare exception to be taught within SRSD is called STOP & DARE, first introduced by de la Paz in 2001. The acronym STOP stands for four different actions to be followed when planning an argumentative written product: (1) suspend judgment (gather arguments for and against a position); (2) take a stand (make up your mind about which side of the argument you want to take); (3) organize ideas (order the arguments according to their strength and persuasiveness); and (4) plan more as you write (keep on planning while finalizing the piece of work). The acronym DARE stands for the four essential main elements of an argumentative essay: (1) develop a topic sentence; (2) add supporting ideas; (3) reject possible arguments for the other side; and (4) end with a conclusion (Kiuhara et al., 2012).

Even though this strategy appears very auspicious and seems to meet all the criteria of an effective technique as outlined in existing meta-analyses (Cook & Bennett, 2014; Datchuk & Kubina, 2015; Gillespie & Graham, 2014; Gillespie et al., 2018; Rogers & Graham, 2008), researchers in only two studies to date have systematically tested the effectiveness of this approach. The first one was conducted by De La Paz and Graham (1997). In their experiment, the authors used STOP & DARE to help three fifth graders with LDs enhance their planning and writing quality. They could demonstrate that all students were able to produce longer and better opinion essays over the course of the intervention. In the second study, Ennis et al. (2013) evaluated STOP & DARE with 16 children with behavior problems from Grades 3 to 6. After the treatment, they wrote significantly more extensive texts and essays of higher quality than the boys and girls in the control group.

As mentioned earlier, composing position papers places especially high demands on students' working memory. When offered instruction on how to produce such texts, most children and youth with severe learning problems feel overwhelmed and discouraged due to the strenuousness that engagement in such an endeavor requires. This suggests the adding of some motivational techniques aimed at increasing the chances of students' willingness to engage in such an arduous intervention. Fortunately, there is ample evidence that adding certain galvanizing components to a treatment serves just this purpose (i.e., Garcia & de Caso, 2006).

A verified technique to increase the motivation to learn is progress monitoring (Förster & Souvignier, 2014; Martin et al., 2002). It is defined as a frequent measurement of targeted skills to evaluate the effectiveness of an intervention and to inform students of their performance development (Shapiro et al., 2011). Another element of this technique is the cognitive reconstruction of maladaptive attributions. Weak learners often do not explain their achievements with their efforts but ascribe them to external reasons such as luck or chance. By contrast, failures tend to be at-

tributed to a lack of ability (Robertson, 2000). As a consequence, the effort to learn is reduced. To alter such unfavorable cognitions, teachers should provide frequent feedback that ascribes success to the students' endeavors and failure to variable reasons, such as "a bad day" (Cameron & Pierce, 1994; Homer et al., 2018). A third well-tried component in this context is self-scoring. Light et al. (1988) showed that students' performance improves when they keep track of their learning advancements. According to Grünke et al. (2017), the combination of the presented motivational modules has a significant impact on the writing performance of students with severe learning problems.

RESEARCH QUESTION

Our aim in this study was to add to the scarce body of literature on the effectiveness of STOP & DARE by testing this intervention with a group of struggling eighth and ninth graders. To increase the likelihood that the participants would engage in the treatment and stick with it until the end, we supplemented the training with three motivational techniques: progress monitoring, adaptive attribution feedback, and self-scoring.

METHOD

Design

We applied a randomized group design with an experimental group (EG) and control group (CG) consisting of a pretest (O1), a posttest (O2), and a follow-up assessment (O3) that we conducted 4 weeks after the posttest (see Figure 1).
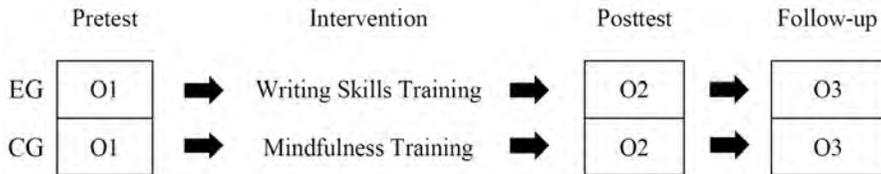

	Pretest		Intervention		Posttest		Follow-up
EG	O1	➡	Writing Skills Training	➡	O2	➡	O3
CG	O1	➡	Mindfulness Training	➡	O2	➡	O3

Figure 1. **Design of the experiment.**

We assigned the participants randomly to the two conditions. To create formally equivalent circumstances and thus ensure a high level of internal validity (Odom et al., 2005, both groups received some kind of intervention that was delivered at the same time, for the same duration, with similar material, and always in the students' familiar classroom environment (see below).

Participants

Our sample consisted of 77 eighth and ninth graders from two schools in a metropolitan region in North Rhine-Westphalia (Germany). One of them was a special school for students with moderate general LDs (Förderschule), and the other

one was a secondary school for less academic students (Hauptschule). All participants were considered to be underperforming in core subject areas. Students' ages varied from 13 to 16 years. We selected this rather old group because boys and girls may hardly be able to acquire complex argumentative essay skills before entering the formal operational stage, according to Piaget (1957), which usually starts at age 12. Of these 77 students, 22 (28.57%) had been diagnosed with an LD by a multi-professional team. This means that they struggled severely to "develop the knowledge, skills, will and self-regulation necessary to succeed in key subject areas" (Grünke & Morrison Cavendish, 2016, p. 1). Thirty-eight of the participants were assigned to the EG, and 39 were assigned to the CG. The random allocation happened separately for students with and without LDs, ensuring that there was an equivalent number of participants with and without LDs in each condition.

Table 1 shows an overview of the demographic characteristics in the EG and the CG. The mean age of the first group was 14 years and 8 months, and the one in the second group was 15 years and 0 months.

Table 1. *Ratio of Different Demographic Variables in Both Groups*

Group	Female students	Migration background	German as second language	LD
EG ($N = 38$)	47.4%	60.9%	39.47%	28.95%
CG ($N = 39$)	48.7%	56.3%	33.33%	28.21%

INSTRUMENTS

We captured argumentative writing performance at all three measurement times through opinion essays that the participants produced in response to writing prompts, which consisted of randomly drawn questions they could relate to (e.g., "Should schools offer cash bonuses for good test scores?"; "Does gym help you perform better in all your classes?"; or "Does technology make us more alone?"). We asked the students to work quietly on their texts, and we did not impose time restrictions on them. However, no one took more than 30 min to complete the task.

Two measures served as dependent variables: *total words written* (TWW) to register productivity and an *argumentative writing rubric* (AWR) to capture text quality. TWW is the most widely used criterion for measuring quantitative progress in the field of text production. It is defined as the number of recognizable words written regardless of spelling or context (Hosp et al., 2016). This option is considered a valid and reliable tool for the quantitative assessment of student writing performance and usually correlates highly with text quality (Grünke et al., 2015).

The AWR was created by the Strategic Education Research Partnership (2003), and it is a scoring guide that describes criteria for the quality of position papers. It was developed as part of an intervention program that has been proven to be very effective (ibid.). The scoring sheet entails four categories (e.g., argumentation, evidence, organization, and language). In relation to each dimension, an argumentative essay can be rated on a four-point scale, ranging from *emerging, developing,*

and *proficient* to *exemplary*. For each grading, a detailed description is provided. The number of points allocated to a writing product can range from 0 (in case not even emerging standards were met) to 4 (in case all the stipulated requirements were fulfilled). Thus, the overall scores could vary between 0 and 16.

A senior researcher and two graduate students of special education (who were blind to the purpose of the study) appraised each text independently using TWW and the AWR as performance indicators. Codification differences between raters were discussed until consensus was reached. In all cases of TWW and AWR assessments, the initial interrater reliability ranged from 90 to 100%.

Procedures

After pretest measurements had been completed and the study sample was divided into the EG and the CG, both groups received one of two types of treatments, which were formally identical despite differing with regard to content. Under each condition, the participants attended 10 training sessions (lasting 45-min each) over a period of three weeks.

Eleven graduate university students served as interventionists. The first author intensively trained them over the course of three 45-min briefing lessons. To further ensure treatment fidelity, we gave the university students a detailed script to follow for each session. In addition, they stayed in constant contact with the first author by phone or e-mail so that she could interfere at any time in case someone strayed from the designated training plan. The interventionists did not always instruct the same group of participants (which each encompassed between 10 and 15 adolescents) but taught them in accordance with a rotation system. We implemented this to minimize the possibility that potential treatment effects could be attributed to the personality of specific university students.

The EG received writing-training based on the STOP & DARE concept (as outlined by De La Paz, 2001), along with a motivation package with the aim of promoting argumentative writing skills. At the beginning of the first session, the interventionists explained about what constitutes an argumentative essay to their students (Step 1 of SRSD). That is, they elucidated that it requires one's viewpoint on a given topic, which must be stated clearly, giving various arguments underpinned by logical reasons. In addition, opposing positions should be expounded and weighed up against one's own notion. Finally, an argumentative essay must always end with a conclusion, in which the author wrapped up the main ideas and left the reader with a well-balanced answer to the question that the essay is addressing.

Subsequently, the interventionists introduced the essay-planning strategy and explained the acronyms of STOP & DARE (Step 2 of SRSD). They did that by going through a checklist that specified all eight action steps of the procedure (see above) and contained boxes for registering if they were observed or not. We used this instrument from then on in every lesson as a self-scoring tool whenever the students worked on writing an essay and reviewed whether they took notice of all the mandatory actions required by STOP & DARE. At the end of the first session, we asked the participants to write an opinion paper about a topic that was randomly selected from a compilation of argumentative writing prompts (the checklist and the set of writing themes can be obtained from the first author upon request).

The second unit began with the interventionists presenting the participants with the results of a progress-monitoring procedure. They showcased line diagrams that visually illustrated the number of TWW that the student essays from the previous day consisted of. The scores were prominently displayed on the first page of workbooks that were handed out to all participants so they could store all the texts they would write over the course of the training. We justified laying the focus on productivity by involving these diagrams because struggling writers usually compose rather short and incomplete texts (MacArthur & Graham, 2016). Encouraging the participants to write more extensive essays is one of the initial provisions when trying to boost their quality (Grünke et al., 2015).

The interventionists then iterated the criteria of a good position paper. They went through the different steps of the strategy and verbally modeled how to execute them using a randomly chosen debate topic from the previously mentioned compilation while thinking aloud (Step 3 of SRSD). Subsequently, the students grouped up in pairs and explained to each other alternately what needs to be done during each phase of the STOP & DARE procedure (Step 4 of SRSD). Finally, we asked them to compose another opinion essay.

The third session started with presenting the updated line diagrams and with praising the students for any improvements in the number of TWW. At this point, we provided motivational feedback on the basis of attribution theory. If one of the participants failed to write a lengthier story than the day before, the respective interventionist suggested that this was due to various reasons (e.g., "Yesterday you seemed very tired; I am sure you will do better today" or "The topic that you wrote about yesterday was probably not cut out for you; you might find the next one more interesting"). In cases where the number of words in an essay had increased, the respective interventionist praised the participant and provided internal attributions for the improvement (e.g. "You tried especially hard and it showed!" or "You have a great talent for coming up with very creative arguments"). If students wrote more than they had ever done before during treatment, the old high score in their workbook was replaced by the new one. After the interventionists recapitulated what makes a good opinion essay, the participants went through the aforementioned checklist in pairs. While doing so, they weighed up different arguments pertaining to topics that were assigned to them. Along the way, the interventionists scaffolded the process as they provided supportive feedback (Step 5 of SRSD). Another argumentative writing assignment marked the end of the lesson.

The rest of the sessions started like the third one, by giving feedback on the line diagrams and explaining to the students that any improvements came from reasons inherent to them, whereas stagnations or decreases in the number of TWW had variable causes. Subsequently, the participants continued to practice finding pro and con arguments for the proposed essay themes, by sequencing them according to their persuasiveness, formulating topic sentences, phrasing balanced conclusions, and so on (Step 6 of SRSD). The interventionists monitored the students' activities and offered help whenever they thought it was needed. Every unit was completed with the participants having composed an opinion essay.

The CG received a formally equivalent mindfulness training according to Kaltwasser (2016). Instead of writing an argumentative text at the end of each session,

the students took note of what they had learned during the intervention that day (a detailed manual on how the treatment was implemented can be obtained from the first author upon request).

<div align="center">RESULTS</div>

Descriptive Analysis

Table 2 contains means (Ms), standard deviations (SDs), and coefficients of variation (CVs) of the EG and the CG at all three measurements for both dependent variables (TWW and AWR). At the pretest condition, we determined almost identical mean values for both subsamples. However, posttest results demonstrated the superiority of the EG over the CG for the two variables. For the number of TWW, the CG showed an increase in performance from O1 to O2 (46.05%), but the improvements in the EG were much greater (368.88%). Both groups achieved slightly lower scores at O3 in comparison to O2 (-8.37% and -7,47%). The AWR scores were almost identical in both groups at pretest. There was a marked increase in the EG from O1 to O2 (122.20%) and a slight decrease from O2 to O3 (-3.17%). In the CG, the scores from O1 to O2 lowered by -3.52%, but and rose from O2 to O3 by 17.12%.

Table 2. *Descriptive Data for TWW and AWR*

		EG (*N* = 38)			CG (*N* = 39)		
		M	*SD*	*CV*	*M*	*SD*	*CV*
TWW	Pretest	29.92	19.99	0.67	30.51	20.51	0.67
	Posttest	140.29	109.35	0.78	44.56	37.80	0.85
	Follow-up	128.55	122.89	0.96	41.23	27.80	0.67
AWR	Pretest	4.55	2.06	0.45	4.54	2.38	0.52
	Posttest	10.11	1.45	0.14	4.38	1.83	0.42
	Follow-up	9.79	1.55	0.16	5.13	1.36	0.27

What is striking is the high variability of the measurements. The CVs are especially high for TWW (ranging from 0.67 to 0.96 in the EG, and from 0.67 to 0.85 in the CG). Interestingly, the CVs for the productivity of the adolescents who learned how to write good argumentative essays rose over time, whereas the CVs for quality decreased. This means that, after the intervention, the students in EG wrote papers that differed very much in length (the CVs increased by 43.28% from O1 to O3), but the level of sophistication became much more similar (the CV decreased by 64.44% from O1 to O3).

Figures 2 and 3 show the listed results for TWW and AWRon the basis of the estimated marginal means in a graphical representation. As can be seen in both graphs of the descriptive data discussed above, the EG clearly outperformed the CG and, moreover, was able to maintain those scores at a fairly high level during the follow-up.
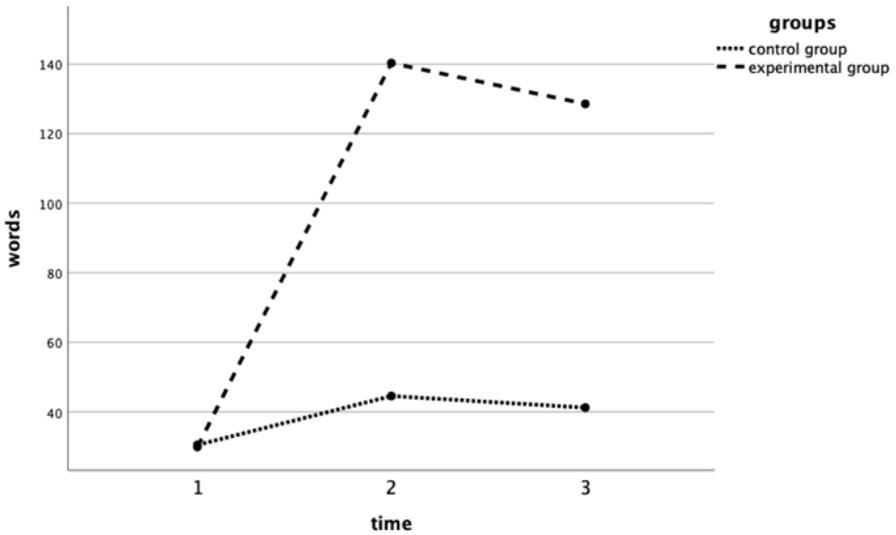
Figure 2. **Graphical representation of the estimated marginal mean (based on the number of written words) over the three measurement times.**
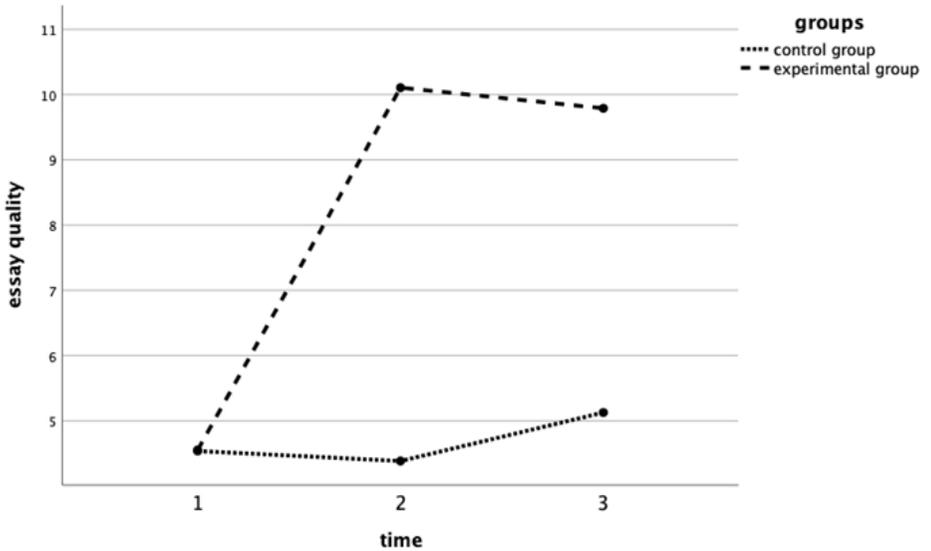

Figure 3. **Graphical representation of the estimated boundary means (based on the AWR evaluation) over the three measuring times.**

Thus, the descriptive statistics indicated a noteworthy increase in the length of the texts, as well as in their quality in the EG, whereas performance of the CG remained more or less the same over time.

Quantitative Analysis

Total words written. In the next step, we executed a mixed analysis of variance (ANOVA) with repeated measurements with post hoc tests for pairwise comparison. Review of the measured data using the Shapiro-Wilk test showed that the TWW values of both groups were not normally distributed ($p < .05$). However, ANOVA is relatively robust to violations of the normal distribution (see Salkind, 2010), which is why we continued to work with this statistical model. With regard to the variance homogeneity, we noticed that at O2 and O3, the values of both groups were not equivalent ($p < .05$). Using the Mauchly test for sphericity, we showed that sphericity was not present ($p < .001$; Field, 2013). Thus, we corrected values using the Greenhouse–Geisser method. The criteria of homogeneity regarding covariance matrices were, in contrast to AWR, not fulfilled ($p = < .001$; ibid.).

A statistically significant interaction effect could be found for time x group ($F[1.55, 116.46] = 22.43, p < .001, \eta^2 = .23$). Moreover, the ANOVA showed a statistically significant impact of time on the EG ($F[1.53, 56.49] = 30.84, p < .001, \eta^2 = .46$) and on the CG ($F[1.47, 55.83] = 5.11, p < .05, \eta^2 = .119$). In addition, we found a significant difference between the two groups ($F[1.75] = 21.36, p < .001, \eta^2 = .22$; see Table 4). The effect size of $\eta^2 = .23$ for the time x group effect can be considered large and is equivalent to a Cohen's d of 1.09.

Table 3. *Interaction and Main Effect for Between Subjects and Within Subjects for TWW Comparing EG and CG*

Factors		*F*	*df1*	*df2*	*p*	η^2
Time x group		22.43	1.55	116.56	< .001	.23
Time	EG	30.84	1.53	56.49	< .001	.46
	CG	5.11	1.47	55.83	< .05	.12
Group		21.36	1	75	< .001	.22

We conducted Bonferroni-adjusted post hoc tests to identify specific differences between groups. Regarding the inner subject factor and the pairwise comparison of measurements for each group, we found that no statistical significance could be maintained for the CG with respect to the difference between O1, O2, and O3. In contrast, differences in the performance-results of the EG between pretest and posttest ($p < .001$) and also between pretest and follow-up ($p < .001$) both proved to be statistically significant. However, the scores at O2 and O3 stayed on a similar level ($p = .824$). Moving on with focusing on the post hoc tests for the intermediate subject factor, using the robust Welch test, EG and CG did not differ in the pretest results ($p = .898$) but grew apart at the posttest ($p < .001$) and the follow-up ($p < .001$). Due to the heteroscedasticity of variances and inhomogeneity regarding covariance matrices, we conducted a robust mixed ANOVA, which yielded similar results for the interaction effect ($F[2, 24.48] = 14.32, p < .01$), main effect for time variable ($F[2, 24.48] = 22.12, p < .001$), and main effect for the intermediate subject factor ($F[1, 29.43] = 20.01, p < .001$; see Table 4).

Table 4. *Pairwise Comparison for the Innersubject Factor of the EG and the CG (TWW)*

Group	Time		M. Diff.	SE	Sig.	d
EG	O1	t2	110.37	16.31	< .001	1.40
		t3	98.63	18.25	< .001	1.12
	O2	t1	110.37	16.31	< .001	1.40
		t3	11.74	10.58	.824	0.10
	O3	t1	98.63	18.35	< .001	1.12
		t2	11.74	10.58	.824	0.10
CG	O1	t2	14.05	5.76	.058	0.46
		t3	10.72	4.29	.051	0.44
	O2	t1	14.05	5.76	.058	0.46
		t3	3.33	3.42	1.00	0.10
	O3	t1	10.72	4.29	.051	0.44
		t2	3.33	3.42	1.00	0.10

Note. EG = experimental group; CG = control group; t1 = pretest; t2 = posttest; t3 = follow-up.

On the basis of the ANOVAs, we determined that there was a significant difference between the groups from the second measurement point onward and that the writing intervention was effective over a longer period of time, which can be observed from the stable data in the follow-up. We did not find such effects for the CG.

Argumentative writing rubric. For ANOVA with repeated measurements, we tested the requirements. With regard to the normal distribution, we noticed that at Measurement Times 2 and 3, the values of both groups were not normally distributed ($p < .05$). The Mauchly test results showed a non given sphericity ($p = < .01$). Thus, we corrected all of the following results using the Greenhouse–Geisser method. The criterion of homogeneity regarding covariance matrices was met ($p = .131$; Field, 2013). Table 5 shows the results of the ANOVA.

Table 5. *Interaction and Main Effect for Between Subjects and Within Subjects for AWR*

Factors		F	df1	df2	p	η^2
Time x group		70.72	1.70	127.72	< .001	.485
Time	EG	146.39	1.53	56.51	< .001	.798
	CG	2.41	1.72	65.45	.105	.060
Group		145.09	1	75	< .001	.659

Note. EG = experimental group; CG = control group.

We found a statistically significant interaction effect between the time and groups (F[1.703, 127.721] = 70.718, $p < .001$, partially $\eta2 = .485$). The effect size of $\eta^2 = .49$ was impressively high and equivalent to a Cohen's d of 1.94. Beyond that, the results did not indicate a statistically significant effect of time on the dependent variable with respect to the CG (F[1.722, 65.451] = 2.412, $p = .105$, $\eta^2 = 0.06$). For the EG, however, we found statistically significant effects of the time variable on writing quality with a remarkable effect size (F[1.525, 56.415] = $p < .001$, $\eta^2 = .798$). Main effects for the intermediate subject factor showed a statistically significant difference between the two groups (F[1.75] = 145.088, $p < .001$, $\eta^2 = .659$). To determine the exact differences, Table 6 shows the pairwise comparison using post hoc tests. As already mentioned, only a small effect occurs between the individual measurement times for the control condition. In the EG, however, it became apparent between which measuring points the statistically significant effects could be found. We noted significant differences between O1 and O2 as well as for O1 and O3. Also, we did not detect significant differences for O2 and O3. Concerning the inhomogeneity of variances, we executed a robust mixed ANOVA (Mair & Wilcox, 2020) showing the same results for the interaction effect (F[2, 41.55] =33.77, $p < .001$), the main effect for time variable (F[2, 41.55] = 39.26, $p < .001$), and the main effect for the intermediate subject factor (F[1, 45.50] = 122.55, $p < .001$).

Table 6. *Pairwise Comparison for the Innersubject Factor of the EG and the CG (AWR)*

Group	Time		M. Diff.	SE	Sig.	d
EG	O1	t2	5.55	.37	< .001	3.12
		t3	5.24	.44	< .001	2.88
	O2	t1	5.55	.37	< .001	3.12
		t3	.32	.26	.705	0.21
	O3	t1	5.24	.44	< .001	2.88
		t2	.32	.26	.705	0.21
CG	O1	t2	.15	.42	1.00	0.08
		t3	.59	.36	.322	0.30
	O2	t1	.15	.42	1.00	0.08
		t3	.74	.29	.042	0.47
	O3	t1	.59	.36	.322	0.30
		t2	.74	.29	.042	0.47

Note. EG = experimental group; CG = control group; t1 = pretest; t2 = posttest; t3 = follow-up.

In summary, the results showed that the students in the EG were able to boost their writing performance greatly, whereas the mindfulness training did not seem to have a momentous impact on the dependent variables.

<div align="center">

DISCUSSION

</div>

Main Findings

Our aim in this study was to add to the scarce reference data on the benefits of the STOP & DARE strategy for secondary students who struggle severely in writing argumentative essays. In our experiment, we evaluated the effects of this technique in combination with a multicomponent motivational intervention, consisting of progress monitoring, adaptive attributional feedback, and self-scoring. The participants in the EG and the CG did not differ in their achievement level before treatment. However, comparing the performance development in both groups, it became evident that the students who were taught the STOP & DARE strategy showed a much greater increase in both text length and quality over the course of the experiment than the ones in the control condition. Although the CG also demonstrated a slight increase in productivity from pre- to posttest, we could not detect a statistically significant difference between the relevant mean scores. The time x group interaction for the indices that were applied to measure text length and quality yielded remarkably high effect sizes of $\eta^2 = .23$ and $\eta^2 = .49$, respectively. Judging by the results of the follow-up measurement, the treatment gains in the EG were comparatively stable. What is especially positive about our findings is the fact that the performance development in the EG was contrasted with the one in a group that received a similar amount of attention and took part in a training that seemed to be just as interesting and stimulating to the participants as the STOP & DARE intervention. Thus, the effects cannot be attributed to the EG being exposed to special circumstances that differed from their usual classroom routine.

The originality and relevance of our study lie in the fact that we were able to document the potency of an underresearched approach that seems to be very suitable for teaching important argumentative writing skills to struggling learners between 13 and 16 years old. According to the model of Grabowski et al. (2014), adolescents of that age should have cleared the developmental hurdle of being able to produce opinion essays of sufficient quality. However, all of our subjects could be considered academically challenged. At the beginning of our study, they composed only very short papers that were weak in content. Upon termination of the treatment, the students who had received the writing skills training produced essays of about 140 words on average after considering most of the criteria of an appealing position paper. The large variance in text length documents how greatly the participants of the EG differed in how much they produced. However, a very positive result is the fact that the effect size for quality was about double that for quantity. What is more, the CV of the number of TWW is significantly higher than the CV of AWR scores. Comparing the values of the follow-up measurement, the ratio is 6:1. This speaks to the assumption that no matter how much the participants wrote, it was of decent quality.

Limitations

Despite the positive results, there are limitations in this study. First of all, the interventionists were not blind to the purpose of the experiment but knew full well that the goal was to test the efficacy of the STOP & DARE strategy. Even though we did not disclose the purpose of the research to our participants, we shared this infor-

<div align="center">

</div>

mation with the graduate university students. Thus, the results might be somewhat biased due to the interventionists' possible preferences for the EG to outperform the CG. This could have subconsciously influenced the commitment and dedication with which they instructed the adolescents in the two groups. However, the time x group interaction effects were so strong that it seems very unlikely that this possible bias was responsible for a noteworthy portion of the differences in the performance development between EG and CG.

Another limiting aspect pertains to the fact that there is no way of telling which of the intervention components was to what extent responsible for the improvements. Our experiment can only provide information on whether there were differences in the performance development between the EG and the CG. As unlikely as it might seem, it is impossible to completely preclude that the training would have been even more effective had we left out one, two, or all three of the motivational elements (e.g., progress monitoring, adaptive attribution feedback, and self-scoring). To shed more light on this question, further and more differentiated studies are warranted.

An additional constraint is the selection of the sample. We chose secondary students of a particular age group. Thus, we cannot draw any conclusions about the benefits that our training would have on another subgroup. Apart from that, even statements about the impact of the intervention on adolescents from the same age bracket are problematic because one must be very cautious when generalizing results from one study to a particular population. What is more, we categorized our participants as underperforming based on the types of schools they attended. The students were allocated to their respective schools because their academic performance was below average. However, using more objective criteria in the selection process and applying standardized tests would have made it easier to replicate our study.

A last limitation relates to the way performance was captured and reflected back to the students. Even though the participants in the EG practiced to evaluate their essays with a checklist that focused very much on the same criteria as the AWR, progress monitoring and the daily feedback at the beginning of sessions 2 to 10 were based on productivity. TWW usually correlates very highly and positively with different quality measures (Gansle et al., 2004). However, in this experiment, the relative variability of TWW increased, whereas the relative variability for AWR decreased from the pre- to posttest. In the case of TWW, the CV in the EG increased by 16.42%, but the CV for the AWR scores decreased by 68.89%. It is out of the question that the focus should be on teaching students to pay heed to the criteria of what constitutes a top-notch essay instead of motivating them to produce a long text. Reflecting on our findings, it might have made more sense to use quality scores when sharing the progress monitoring results with the students instead of informing them how many words they jotted down the day before. For prospective projects, we suggest giving learners frequent feedback on how well they wrote as opposed to letting them know how much they wrote.

Outlook

The results of this study are in line with the two previous experiments on the benefits of STOP & DARE and with other research focusing on similar strategies.

Previous authors have underlined the assumption that the challenges that under-performing students face when composing an argumentative paper can be remedied with simple, easy-to-use, and time-efficient tools. We implemented our project under realistic conditions of everyday life in school. Therefore, it can easily be tailored to the specifics of a given classroom situation. The costs for the materials were very low, and the interventionists needed comparatively little instruction in preparation for carrying out the treatment. Given the fact that the ability to compose acceptable argumentative essays is vital to one's aptitude to think critically and one's cognitive development in general, this important competency should be pulled out of the dusty corners of the classroom. Educators all too often focus on reading and spelling, while neglecting expressive writing, especially in the area of relatively ambitious genres such as argumentative essays (Grünke & Leonard-Zabel, 2015). An easy-to-teach strategy like STOP & DARE in combination with different motivational elements can come to the rescue. It remains to be hoped that future studies will add to the knowledge base concerning the differential benefits of this approach and that it will be widely disseminated to improve the chances of struggling students in education systems and the labor markets around the world.

References

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

Cameron, J., & Pierce, W. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *American Educational Research Association, 64*, 363–423. https://doi.org/10.3102/00346543064003363

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. (2015). Effects of sentence-combining instruction and frequency building to a performance criterion on adolescents with difficulty constructing sentences. *Journal of Evidence-Based Practices for Schools, 14*, 160–185.

De La Paz, S. (2001). Stop & Dare: A persuasive writing strategy. *Interventions in school and clinic, 36*, 234–243. https://doi.org/10.1177/105345120103600409

De La Paz, S., & Graham, S. (1997). Strategy instruction in planning: effects on the writing performance and behavior of students with learning difficulties. *Exceptional Children, 63*, 167–181. https://doi.org/10.1177/001440299706300202

Ennis, R. P., Jolivette, K., & Boden, L. J. (2013). STOP and DARE: Self-regulated strategy development for persuasive writing with elementary students with E/BD in a residential facility. *Education and Treatment of Children, 26*, 81–99. https://doi.org/10.1353/etc.2013.0026

Fayol, M. L., Alamargot, D., & Berninger, V. W. (2012). From cave writers to elite scribes to professional writers to universal writers, translating is fundamental to writing. In M. L. Fayol, D. Alamargot, & V. W. Berninger (Eds.), *Translation of thought to written text while composing* (pp. 3–14). New York, NY: Taylor & Francis. https://doi.org/10.4324/9780203141434

Field, A. (2013). *Discovering statistics using. IBM SPSS Statistics* (4th ed.). London, UK: SAGE Publications.

Förster, N., & Souvignier, E. (2014). Learning progress assessment and goal setting: Effects on reading achievement, reading motivation and reading self-concept. *Learning and Instruction, 32*, 91–100. https://doi.org/10.1016/j.learninstruc.2014.02.002

Gansle, K. A., Noell, G. H., Van Der Heyden, A. M., Slider, N.J., Hoffpauir, L. D., Whitmarsh, E. L., et al. (2004). An examination of the criterion validity and sensitivity to brief interventions of alternate curriculum-based measures of writing skill. *Psychology in the Schools, 41*, 291–300. https://doi.org/10.1002/pits.10166

García, J. N., & de Caso, A. M. M. (2006). Changes in writing self-efficacy and writing products and processes through specific training in the self-efficacy beliefs of students the learning disabled. *Learning Disabilities. A Contemporary Journal, 4*, 1–27.

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. https://doi.org/10.18666/LDMJ-2018-V23-I2-8990

Grabowski, J., Becker-Mrotzek, M., Knopp, M., Jost, J., & Weinzierl, C. (2014). Comparing and combining different approaches to the assessment of text quality. In D. Knorr, C. Heine, & J. Engberg (Eds.), *Methods in writing process research* (pp. 147–165). Frankfurt/M., Germany: Lang.

Graham, S. (2006). Strategy instruction and the teaching of writing: A meta- analysis. In C. MacArthur, S. Graham, & F. Fitzgerald (Eds.), *Handbook of writing research* (pp. 187–207). New York, NY: Guilford.

Graham S. (2007). Research on writing development, practice, instruction, and assessment. *Reading and Writing, 21*, 1–2. http://doi.org/10.1007/s11145-007-9069-7

Graham, S., & Harris, K. R. (2000). The role of self-regulation and transcription skills in writing and writing development. *Educational Psychologist, 35*, 3–12. https://doi.org/10.1207/S15326985EP3501_2

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*, 19–33. https://doi.org/10.1177/00224669050390010301

Graham, S., Harris, K. R., & McKeown, D. (2013). The writing of students with LD and a meta-analysis of SRSD writing intervention studies: Redux. In H. L. Swanson, K. Harris, & S. Graham (Eds.), *The handbook of learning disabilities* (2nd ed., pp. 405–438). New York, NY: Guilford.

Graham, S., Harris, K. R., & Santangelo, T. (2015). Research-based writing practices and the common core. Meta-analysis and meta-synthesis. *The Elementary School-Journal, 115*, 498–522. https://doi.org/10.1086/681964

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

Graham, S., & Perin, D. (2007a). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools – A report to Carnegie Corporation of New York.* Washington, DC: Alliance for Excellent Education.

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

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

Grünke, M., & Morrison Cavendish, W. (2016). Learning disabilities around the globe: Making sense of the heterogeneity of the different viewpoints. *Learning Disabilities: A Contemporary Journal, 14*, 1–8.

Grünke, M., Sperling, M., & Burke, M. D. (2017). The impact of explicit timing, immediate feedback, and positive reinforcement on the writing outcomes of academically and behaviorally struggling fifth-grade students. *Insights into Learning Disabilities, 14*, 135–153.

Harris, K. R., Graham, S., & Friedlander, B. (2013). Bring powerful writing strategies into your classroom! *Reading Teacher, 66*, 538–542. https://doi.org/10.1002/TRTR.1156

Harris, K. R., Graham, S., Mason, L. H., & Friedlander, B. (2008*). Powerful writing strategies for all students*. Baltimore, MD: Paul H. Brookes Publishing Co.

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). Hillsdale, NY: Lawrence-Erlbaum.

Homer, R., Hew, K., & Tan, C. (2018). Comparing digital badges-and points with classroom token systems: Effects on elementary school ESL students' classroom behavior and English learning. *Journal of Educational Technology & Society*, *21*, 137–151.

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2016). *The ABCs of CBM: A Practical Guide to Curriculum-Based Measurement (vol. 2)*. New York, NY: Guilford.

Kaltwasser, V. (2016). *Praxisbuch Achtsamkeit in der Schule* [Practice book mindfulness in school]. Weinheim, Germany: Beltz.

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), Rochester, Minnesota. *Pediatrics*, *123*, 1306–1313. https://doi.org/10.1542/peds.2008-2098

Kiuhara, S. A., O'Neill, R. E., Hawken, L. S., & Graham, S. (2012). The effectiveness of teaching 10th-grade students STOP, AIMS, and DARE for planning and drafting persuasive text. *Exceptional Children, 78*, 335–355. https://doi.org/10.1177/001440291207800305

Light, L. C., McKeachie, W. J., & Lin, Y. G. (1988). Self-scoring: A self-monitoring procedure. *Teaching of Psychology, 15*, 145–147. https://doi.org/10.1207/s15328023top1503_11

MacArthur, C. A., & Graham, S. (2016). Writing research from a cognitive perspective. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed., pp. 24–40). New York, NY: Guilford.

Mair, P., Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*, *52*, 464–488. https://doi.org/10.3758/s13428-019-01246-w

Martin, J. J., Kulinna, P. H., & Cothran, D. (2002). Motivating students through assessment. *Journal of Physical Education, Recreation and Dance, 73*, 18–30. https://doi.org/10.1080/07303084.2002.10608324

Mason, L. H., & Shriner, J. G. (2008). Self-regulated strategy development instruction for writing an opinion essay: Effects for six students with emotional/behavior disorders. *Reading and Writing: An Interdisciplinary Journal, 21*, 71–93. https://doi.org/10.1007/s11145-007-9065-y

Mayer, C. (2007). What really matters in the early literacy development of deaf children. *Journal of Deaf Studies and Deaf Education, 12*, 411–431. https://doi.org/10.1093/deafed/enm020

O'Brien, D., Beach, R., & Scharber, C. (2007) "Struggling" middle schoolers: Engagement and literate competence in a reading writing intervention class. *Reading Psychology*, *28*, 51–73. https://doi.org/10.1080/02702710601115463

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children*, *71*, 137–148. https://doi.org/10.1177/001440290507100201

Piaget, J. (1957). *Construction of reality in the child*. London, UK: Routledge & Kegan Paul.

Robertson, J. S. (2000). Is attribution training a worthwhile classroom intervention for K-12 students with learning difficulties? *Educational Psychology Review, 12*, 111–134. https://doi.org/10.1023/A:1009089118008

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

Salkind, N. J. (2010). *Encyclopedia of research design (vol. 2).* Thousand Oaks, CA: SAGE Publications. http://dx.doi.org/10.4135/9781412961288

Santangelo, T. (2014). Why is writing so difficult for students with learning disabilities? A narrative review to inform the design of effective instruction. *Learning Disabilities: A Contemporary Journal, 12,* 5–10.

Shapiro, E. S., Zigmond, N., Wallace, T., & Marston, D. (2011). *Implementing response to intervention: Tools, outcome, and implications.* New York, NY: Guilford.

Strategic Education Research Partnership (2003, June 25). *Argumentative writing rubric.* https://wordgen.serpmedia.org/assets/wordgen_argumentative_writing_rubric.pdf

Troia, G. A. (2009). *Instruction and assessment for struggling writers.* New York, NY: Guilford.